Numerical stability of generalized entropies

György Steinbrecher
University of Craiova, A. I. Cuza 13, 200585 Craiova, Romania.
Email: gyorgy.steinbrecher@gmail.com

Giorgio Sonnino
Department of Theoretical Physics and Mathematics, Université Libre de Bruxelles (ULB), Campus Plain CP 231, Boulevard de Triomphe, 1050 Brussels, Belgium &
Royal Military School (RMS)
Av. de la Renaissance 30, 1000 Brussels - Belgium.
E-mail: gsonnino@ulb.ac.be

May 20, 2016

Abstract

In many applications, the probability density function is subject to experimental errors. In this work the continuos dependence of a class of generalized entropies on the experimental errors is studied. This class includes the C. Shannon, C. Tsallis, A. Rényi and generalized Rényi entropies. By using the connection between Rényi or Tsallis entropies, and the *distance* in a the Lebesgue functional spaces, we introduce a further extensive generalizations of the Rényi entropy. In this work we suppose that the experimental error is measured by some generalized L^p distance. In line with the methodology normally used for treating the so called *ill-posed problems*, auxiliary stabilizing conditions are determined, such that small - in the sense of L^p metric - experimental errors provoke small variations of the classical and generalized entropies. These stabilizing conditions are formulated in terms of L^p metric in a class of generalized L^p spaces of functions. Shannon's entropy requires, however, more restrictive stabilizing conditions.

PACS: 05.70.Ln, 02.50.Cw, 02.30.Cj, 07.05.Kf, 02.30.Zz

1 Introduction

The generalizations of the Boltzmann-Shannon entropy (BSE) [1], the Rényi entropy (RE) [2]-[5] and the Tsallis entropy (TE) [6]-[9] were introduced, independently, by an axiomatic respectively physical approach. These entropies are

algebraically related to the same norm, or *distance functional*, in suitable L^p spaces [10]. There are many applications of this class of generalized entropies [8], [9], [11]-[15]. The problem of mathematical naturalness of the generalized entropies, in the sense of category theory, is treated in [16]. Since in the generic cases, the classical and generalized entropies, are functionals defined in infinite-dimensional L^p spaces, where p is exactly the parameter in the RE and TE (see [10] and references therein), the first problem is related to the correct domain of definition, the finiteness of the defining functionals. Here, we shall use the general, Lebesgue L^p space formalism [10], [16].

In applications the probability density function (PDF) is known only approximately, so the second, practical, problem concerns the numerical stability of the generalized and the classical entropies when the PDF is known only approximately. This stability problem is the main subject of the our study. This problem was studied also in relation to the convergence in Central Limit Theorem in [17]. Since the PDF is always integrable, one of the natural measure of the approximation of PDF is the L^1 *distance* between the exact and approximate PDFs. This *distance* is always finite. In the case of discrete probability distribution, the numerical stability is studied in [18], where numerical stability is formulated in terms of l^1 distance for the input-data, and of *relative error of the entropy* for the output-data. We are pushed to extend the concepts of *finiteness* and *numerical stability* by using the general formalism of the probability theory in general measure spaces [19]-[21], that include the l^1 space. Generally, when the only informations is related to the L^1 norm, we have no control on the other L^p distances, both for RE and TE. Both TE and RE (in the sequel, referred to as RTE) may be expressed in terms of L^p-distances of the PDFs. However in general, despite the L^1-convergence of the approximate PDF sequence, different L^p-convergences (and topologies) are not equivalent [19]-[21], in contradistinction to finite dimensional case where all of the l^p norms are equivalent. It is clear that in the generic infinite-dimensional case, the continuity of entropies in the case of approximation of PDFs cannot be deduced without evoking extra stabilizing conditions, conditions that make the problem similar to finite dimensional case: in the subset defined by stabilization conditions, the convergence say in L^1 norm implies convergence in a family of L^p norm (see below). In our work we follow the methods adopted for treating the so called *ill-posed problem* (in the sense of J. Hadamard). This method is used in the field of strong interactions in physics not only for establishing numerical methods [22], [23], but also for formulating the problem in more rigorous terms [24]-[27]. Likely, this method can be reformulated appropriately to tackle and solve our problem. In this work we shall firstly find auxiliary *stabilizing conditions* e.g., by imposing that RTE is finite for a particular value of the parameter p. Secondly, we shall prove that, by imposing these stabilizing conditions, for a large range of the parameter p, the *absolute error* for the RTE is controlled by the L^1 error in the input-PDFs. The reason to use the absolute error instead of the relative error is discussed in [28]. This choice is also justified in a counter example shown and discussed in the Subsection 2.2.2. Indeed, it is possible to have very small relative errors in TE even when TE diverges. By using the absolute error as a

measure of numerical stability, it follows that the quantities that are related by continuos functions, like RTE, as well as the corresponding L^p distances, must have similar numerical stability properties. Note that special attention should be addressed to the limit case of BSE: the conditions that stabilize RTE are not sufficient to stabilize BSE, more stabilizing conditions are necessary and in the stability proof the powerful functional analytic methods used previously in high-energy physics are involved [24]-[27].

The classical definitions of the quantity of information, or of the degree of randomness, at least in the case of discrete (atomic) phase space, tacitly assume the invariance of the quantity of information, or entropy, under measure preserving groups that in classical simple cases are the permutations [1]-[3], [6] -[9]. *This symmetry assumption applies to the case of more abstract definitions i.e. the previous definitions of the entropy are invariant under the group of measure preserving transformations. This reflects our complete lack of knowledge on the possible dynamical behavior [14]. Larger symmetry group means less information.* Nevertheless, for real physical systems, for instance when the phase space is a direct product of subspaces having completely different physical interpretation and different mathematical structures (e.g., the case when the total system splits into *driving* and *driven* system [10]), we have already an information. In such a situation, the general measure preserving transformations has no physical interpretation and should not be considered as a *true*, we mean *physical*, symmetry [16]. Consequently, we need a new generalization of the entropy able to take into account this direct product structure. The first step in this direction can be found in a previous work cited in Ref. [10] where a new class of extensive generalization of Rényi's entropy is introduced to analyze the case of a phase space, which may be split as a direct product of two sub-spaces (note that the approach illustrated in this work applies equally to the case where the phase-space may be split in an arbitrary, but finite, number of direct products of sub-spaces, like in data with tensor structures).

In this work our aim is to study the stability properties of this class of entropies. As in the previous work [10], our guiding line is to re-interpret the generalized entropy in terms of distance in some generalization of classical [19]-[29] or anisotropic L^p spaces [30].

The main results is the possibility to stabilize RTE or GRE for a large domains of the defining parameters by imposing the finiteness of RTE or GRE for some special values of parameter. An interesting result is that for the stabilization of BSE we need more stabilizing conditions, the BSE is more sensitive compared to general RTE.

The properties of L^p spaces, and associated "distances", will be used for setting the numerical stability properties when auxiliary *stabilizing conditions* are imposed. In this work the stabilizing conditions are expressed as *finiteness* of some classical or generalized (anisotropic [30]) Lebesgue space distances. In some particular cases, these stabilizing conditions may be found by studying directly the partial differential equations for the PDFs. This allows obtaining bounds on the Sobolev space norm and, by using the embedding theorems, getting information on the L^p norms [30], [31]. Another possibility is the study

of the *heavy tail index [33],* [34], [35].

The main results is the possibility to stabilize RTE or GRE for a large domains of the defining parameters by imposing the finiteness of RTE or GRE for some special values of parameter. An interesting result is that for the stabilization of BSE we need more stabilizing conditions, the BSE is more sensitive compared to general RTE.

The work is organized as follows. In Section 2, Subsection 2.1, we relate the RTE to a norm or pseudo-norm in a suitable defined L^p spaces. In the Subsection 2.2, the domain of definition and the problem of continuity of the Shannon, Rényi and Tsallis entropies are studied, when the PDF is approximated in the sense of distance in L^p spaces, with particular emphasis on the natural (from probabilistic point of view) L^1 distance. In particular, we discuss about different concepts of stability (*continuity*, in mathematical terms) of the entropies when the PDF is known only approximately. We find *sufficient* stabilizing conditions for generic RTE. It is also proven (by counter examples) that in the case of BSE, when a sequence of PDF ρ_n converges in the L^1 norm to the PDFρ, extra stabilizing conditions are necessary to ensure the continuity of BSE. We determine the stability of BSE under suitable stabilizing conditions. We emphasize the central role of the functional $Z_q[\rho]$ (see below, [16]) that appears in the both definition of RTE and BSE. The special analytic properties of this functional as well as its relation to the distances in L^p spaces are used in the proofs.

In Section 3, we introduce a further, natural, extension of the Generalized Rényi Entropy (GRE), studied in the previous work ref.([10], [13], [16]), its general properties are studied and some physical applications are presented. In the Section 4 the finiteness and continuity results from Section 2 are extended to the most general case of GRE assuming suitable stabilizing conditions.

We emphasize that the stabilizing conditions are expressed in the terms of measure theoretic concepts. Hence, the results can equally be applied to both continuos and discrete probability distributions, on a general measured space. No continuity or differentiable properties of the PDF are assumed.

Mathematical details can be found in the Appendix 6. In Subsection6.1, the powerful analytic extrapolation method is exposed. This method is necessary to prove the stability of BSE. In Subsection 6.2 and In Subsection 6.3 we found details on the proofs of counter examples and we expose the properties of a class of metric vector spaces, related to the definition and properties of GRE, respectively. In Subsection 6.4, the logarithmic convexity of a class of functionals related to RTE and GRE is presented. These functionals are used in the main proofs of the bound and the numerical stability of the RTE and GRE. For easy reference, since our proof will be generalized to the study of the GRE, some of the known logarithmic convexity properties of the classical L^p norms [21] are demonstrated. The long proofs of stability of the classical Shannon-Boltzmann entropy and of GRE are also presented in detail.

2 Finiteness and continuity aspects of the Shannon, Rényi and Tsallis entropies

2.1 Definitions of generalized entropies in term of norms and pseudo norms

Let's consider a standard measure space $(\Omega, \mathcal{A}, m)$ where Ω is the phase space of a natural system (usually $\mathbf{R}^n$, or in simplest probability models, some finite or denumerable set, or the space of all Brownian trajectories), $\mathcal{A}$ is some $\sigma-$algebra generated by a family-subsets of Ω (that are usually viewed as the collection of the observable events in a probability space), and m is a σ-finite measure defined on $\mathcal{A}$. In this framework the Rényi divergence is expressed by Rényi entropy [16].

In the simplest cases of discrete (atomic) probability spaces, the measure (not necessary finite) m is a counting measure. Note that in this simple case the counting measure is invariant under the group of permutations on Ω. In general, the measure m is chosen to be invariant under some symmetry group of the physical system under study, but there are simple examples for measures with trivial measure preserving group [16]. This formalism allows to consider mixture the discrete and continuos distributions in many variables, and to express the Rényi divergence by RE [16]. In the sequel, we consider only probability measures defined on $(\Omega, \mathcal{A})$ that are continuos with respect to some fixed measure m, that means that for $A \in \mathcal{A}$ the probability $p(A)$ can be represented by probability density function (PDF) ρ as follows

$$p(A) = \int\limits_A \rho(x) dm(x);\ \ A \subset \Omega \tag{1}$$

$$\rho(x) \geq 0;\ \ \int\limits_\Omega \rho(x) dm(x) = 1 \tag{2}$$

or in shorthand notation $dp(x) = \rho(x)dm(x)$. In this framework, the classical BSE is given by

$$S_{cl}[\rho] = -\int\limits_\Omega \log\left[\rho(x)\right] \rho(x) dm(x) \tag{3}$$

The next subsequent generalizations, Rényi's [2]-[5] and Tsallis' [6]-[9], [7] entropies can be reformulated in the terms of distances [10] in the $L^p(\Omega, dm)$ spaces of the distribution functions as follows. They correspond to the norms $\|\rho\|_p$ [20] for $p \geq 1$ and the pseudo-norm $N_p[\rho]$ [19], [29], [10] for $0 < p < 1$,

respectively:

$$\|\rho\|_{p,m} = \left[\int\limits_{\Omega} [\rho(x)]^p \, dm(x) \right]^{\frac{1}{p}} \; ; \; p \geq 1 \tag{4}$$

$$N_{p,m}[\rho] = \int\limits_{\Omega} [\rho(x)]^p \, dm(x); \; 0 < p \leq 1 \tag{5}$$

Details can be found in Ref. [10], Section III. In order to have a unified treatment for both cases $p \lesseqgtr 1$, we introduce the following basic *condensed* notations:

$$i(p) := 1/p \;\; for \; p \geq 1; \; i(p) := 1 \; for \; p \leq 1 \tag{6}$$

$$D_{p,m}[f] := \left[\int\limits_{\Omega} |f(x)|^p \, dm(x) \right]^{i(p)} \tag{7}$$

$$Z_{w,m}[f] := \int\limits_{\Omega} |f(x)|^w \, dm(x); \; \mathrm{Re}(w) > 0 \tag{8}$$

Remark 1 *The function $D_{p,m}[f]$ is a generalized distance from the "point" f (in the function space) to the origin. Note that particular cases of $D_{p,m}[f]$ are $\|f\|_{p,m}$ ($p \geq 1$) and $N_{p,m}[\rho]$ (for $0 < p \leq 1$) defined in Eqs.(4, 5). The functional $f \to D_{p,m}[f]$ has following basic metric properties:*

$$D_{p,m}[f_1 + f_2] \leq D_{p,m}[f_1] + D_{p,m}[f_2] \tag{9}$$

$$|D_{p,m}[f_1] - D_{p,m}[f_2]| \leq D_{p,m}[f_1 - f_2] \tag{10}$$

$$D_{p,m}[\alpha f] = |\alpha|^{s(p)} \; D_{p,m}[f] \; ; \; \alpha \in \mathbb{R} \tag{11}$$

$$s(p) = p i(p) \tag{12}$$

For $p \geq 1$ the functional $D_{p,m}[f] = \|f\|_{p,m}$ (i.e., the classical L^p norm), and for $0 < p \leq 1$ the functional $D_{p,m}[f] = N_{p,m}[f]$ (i.e., the "exotic" L^p pseudo norm from Refs.([19]), [29]). The "precision of approximation ρ_n of the true PDF ρ " is quantified by $D_{p.m}[\rho - \rho_n]$ for some $p > 0$. A possible natural choice is $p = 1$, because $D_{1,m}[\rho - \rho_n] = \int_{\Omega} |\rho - \rho_n| \, dm$ is at least well defined due to Eq.(2).

The corresponding generalized entropies proposed by Rényi [2]-[5] $S_{R,q,m}$ and by C. Tsallis [6]-[9] $S_{T,q,m}$, are given respectively by

$$S_{R,q,m}[\rho] = \frac{1}{1-q} \log Z_{q,m}[\rho] \tag{13}$$

$$S_{R,q,m}[\rho] = \frac{1}{(1-q)i(q)} \log D_{q,m}[\rho] \tag{14}$$

$$S_{T,q,m}[\rho] = \frac{1}{1-q} \left[1 - Z_{q,m}[\rho]\right] = \frac{1}{1-q} \left\{ 1 - [D_{q,m}[\rho]]^{1/i(q)} \right\} \tag{15}$$

From Eqs.(3, 8) results

$$S_{cl}[\rho] = -\frac{d}{dw} Z_{w,m}(\rho)|_{w=1} \tag{16}$$

when it exists. The functionals $S_{T,q,m}[\rho]$, $S_{T,q,m}[\rho]$, $D_{q,m}[\rho]$ and $Z_{q,m}[\rho]$ are algebraically related so they contain exactly the same amount of information. Functional $Z_{q,m}[\rho]$ is the best candidate for including the entropy related functionals in a powerful formalism of category theory [16]. This functional possesses also the remarkable analytic and log-convex properties (see below). In the following proof, we shall use successively the representation of the entropies by functionals $Z_{q,m}[\rho]$ and the distance function $D_{q,m}[\rho]$. From the previous definitions we get the following result:

Remark 2 *Let* $\rho_n, \rho \in L^p(\Omega, dm) \cap L^1(\Omega, dm)$, $p \neq 1$ *with* $\int_\Omega \rho_n(x)dm(x) = \int_\Omega \rho(x)dm(x) = 1$. *The following convergencies are equivalent:*

$$\begin{aligned} &D_{q,m}[\rho_n] \underset{n\to\infty}{\to} D_{q,m}[\rho] \\ &Z_{q,m}[\rho_n] \underset{n\to\infty}{\to} Z_{q,m}[\rho] \\ &S_{T,q,m}[\rho_n] \underset{n\to\infty}{\to} S_{T,q,m}[\rho] \end{aligned}$$

If in addition $D_{q,m}[\rho] > 0$ *then the previous convergences are equivalent to the statement* $S_{R,q,m}[\rho_n] \underset{n\to\infty}{\to} S_{R,q,m}[\rho]$.

Remark 3 *Note that when* Ω *is a finite or denumerable set, if we denote by* p_k *the probabilities of element* $x_k \in \Omega$, *with the measure* m *denoting the counting measure (equal to number of elements in a subset) on the space* Ω, *then from the previous Eqs.(4-14) we get the original definitions Refs. [2]-[5], [6]-[9]*

$$S_{R,q}[\rho] = \frac{1}{1-q} \log \sum_k p_k^q \tag{17}$$

$$S_{T,q}[\rho] = \frac{1}{1-q} \left[\sum_k p_k^q - 1 \right] \tag{18}$$

2.2 Domain of definition (finiteness) and continuity of the Tsallis, Rényi and Shannon entropies.

2.2.1 Generic case: Tsallis and Rényi entropies

Domain of finiteness of RTE's. In the study of the stability (or continuity) properties we consider the general case, when the measure $m(\Omega)$ is not necessary finite and its support does not reduce, necessarily, to a denumerable set. By simple inspection of the Eqs.(8, 13-15) we conclude that the Rényi and Tsallis entropy (RTE) with same q as well as the functionals $D_{q,m}$ and $Z_{q,m}$, are finite

or infinite, simultaneously. The problem related to the domain of definition and continuity of the generalized entropies, when some approximations of the probability density are used, is reduced to the problem of finiteness of continuity of the norm Eq.(4), pseudo norms Eq.(5) or the functional $Z_w(\rho)$ from Eq.(8), respectively. Note that in the case of general measure space, when the measure m is neither atomic nor probabilistic, for instance when $\Omega = \mathbb{R}$ and $dm(x) = dx$, when $q_1 \neq q_2$ the norms $\|\rho\|_{q_1}$, $\|\rho\|_{q_2}$ as well as the pseudo norms $N_{q_1}[\rho]$, $N_{q_2}[\rho]$, are not equivalent (the finiteness or convergence of a sequence ρ_n in the norm $\|\rho_n\|_{q_1}$ is not related to finiteness or convergence in the norm $\|\rho_n\|_{q_2}$ [20], [19] and there is no general inequality relating them). Consider the following example.

Example 4 *Let $\Omega = \mathbb{R}$, $dm(x) = dx$ and consider the PDF $\rho_1(x)$, such that $\int\limits_{\mathbb{R}} \rho_1(x)dx = \|\rho_1\|_1 = 1$ and even more $\int\limits_{\mathbb{R}} |\rho_1(x)\log\rho_1(x)|\, dx < \infty$, $\int\limits_{\mathbb{R}} |\rho_1(x)|^r dx < \infty$ for all $r > 0$. Denote $\rho_\lambda(x) := \lambda\rho_1(\lambda x)$. Despite $\|\rho_\lambda\|_1 = 1$ observe that when $\lambda \to \infty$ for $p > 1$ the RTE diverges, for $0 < p < 1$ the RE diverges. In the limit $\lambda \to 0$ for $p < 1$ the RTE diverges and for $p > 1$ the RE diverges. The BSE of $\rho_\lambda(x)$ diverges both in limits $\lambda \to \infty$ and $\lambda \to 0$.*

It follows that at first sight the answer to the previous questions is tautological. But if we add the physical conditions Eqs (2) we can obtain more information on the finiteness and convergence. The following lemma comes from the well known results [21], when $p > 1$. For easy reference we give an elementary proof, which will be generalized to the study of GRE (Generalized Rényi entropies).

Lemma 5 *Consider the measure space $(\Omega, \mathcal{A}, m)$ and let $f \in L^1(\Omega, dm) \cap L^s(\Omega, dm)$, $s \neq 1$, $s > 0$ with the property*

$$\int\limits_{\Omega} |f(x)|\, dm(x) \leq a_1 \tag{19}$$

$$\int\limits_{\Omega} |f(x)|^s\, dm(x) \leq a_s \tag{20}$$

Then for all r in the range

$$\min\ (1, s) < r < \max\ (1, s)$$

we have also the bound

$$\int\limits_{\Omega} |f(x)|^r\, dm(x) \leq a_1^{\frac{s-r}{s-1}} a_s^{\frac{r-1}{s-1}} \tag{21}$$

This Lemma is a particular case of the Theorem 29, from the Appendix 6.4. Before studying the continuity (also called stability) of the generalized entropies with respect to small variations of the input PDF, firstly we should discuss the

problem of the finiteness of the generalized entropies. The singular case of classical entropy will be discussed later. From the previous Lemma 5 results the following

Proposition 6 *Let $p \neq 1$ and $\min\ (1,p) < r < \max\ (1,p)$. For the PDF ρ obeying Eq.(2), $\rho \in L^1(\Omega, dm) \cap L^p(\Omega, dm)$, and $\int\limits_{\Omega} [\rho(x)]^p dm(x) \leq a_p$ we have*

$$\int\limits_{\Omega} [\rho(x)]^r dm(x) \leq a_p^{\frac{r-1}{p-1}} \tag{22}$$

Consequently if the norm $\|\rho\|_p$ is finite for some fixed index index $p > 1$, then $\|\rho\|_r$, $S_{R,r}[\rho]$, $S_{T,r}[\rho]$ remain finite for all r in the range $1 < r \leq p$. If for some fixed p, with $0 < p < 1$ the pseudonorm $N_p[\rho]$ is finite, then for all r in the range $p \leq r < 1$ also $N_r[\rho]$, $S_{R,r}[\rho]$, $S_{T,r}[\rho]$ remain finite.
Proof. *By setting in Lemma 5 $f(x) = \rho(x)$ from the normalization condition Eq.(2) results $a_1 = 1$. Eq.(22) results directly from Eq.(21). The finiteness of entropies results from Eqs.(8, 13, 15).* ■

When $\Omega = \mathbb{R}$ and $dm(x) := dx$, the pseudo-norm $N_p[\rho]$, and the generalized entropies $S_{R,p}[\rho]$, $S_{T,p}[\rho]$ are divergent for low values of $p < p_0 < 1$ (or, in practical estimations, they have large fluctuations), we may argue that the PDF has a heavy tail: $\rho(x) \underset{|x|\to\infty}{\asymp} |x|^{-1/p_0}$, like in models of stochasticity-induced instability [33], [35], [34]. If there exists some p_1 such that for $p > p_1 > 1$ the entropies, and both norm and $S_{R,p}[\rho]$, $S_{T,p}[\rho]$ are divergent, this suggests that the PDF has an integrable singularity of the type $|x - x_0|^{-1/p_1}$, for instance in the models of noise driven intermittency [36]. Similarly, in the case of PDF defined in higher dimensional, anisotropic space the maximal domain of definition of generalized Rényi entropies [10] is related to more complicated singularity structure and asymptotic behavior of the multivariate PDF.

In the continuation let us suppose to have an exact PDF $\rho \in L^p(\Omega, dm) \cap L^1(\Omega, dm)$, with $p \neq 1$, which is approximated by an approximant sequence $\rho_n \in L^p(\Omega, dm) \cap L^1(\Omega, dm)$, and ρ, ρ_n satisfy the Eq.(2). We also suppose that the "true" limit PDF ρ exists, so ρ_n is a Cauchy sequence, in L^1.

Remark 7 *Our approach on the stability problem is different from Ref.[18], where no stabilizing conditions are imposed. There are simple counter examples 2.2.2 of sequences of probability distributions on $\mathbb{N}$ that are convergent in l^1, consequently bounded and convergent in l^p $(p > 1)$ norm but the BSE diverges, despite it is Lesche stable (see below). We mention that in the case when we restrict ourselves to discrete distributions, from all of the convergence results based on stabilizing conditions it follows also the Lesche stability. The counter example from 2.2.2 proves that there exist sequences of PDF's ρ_n that are convergent in all the spaces $L^p([0, 1/2], dx)$ with $0 < p \leq 1$, nevertheless the BSE diverges.*

Continuity of RTE's when the PDF is approximated in L^1 norm. Suppose that for some $p \neq 1$ we have L^p bounds

$$\int_\Omega |\rho_n(x)|^p \, dm(x) \leq b_1$$

$$\int_\Omega |\rho(x)|^p \, dm(x) \leq b_2$$

For instance, such kind of bounds could be obtained by the technique used in the study of the heavy tail phenomena in random affine processes [34], [35]. The quality of the approximation is quantified in the L^1 norm $\|\rho_n - \rho\|_{L_1}$. By using Eq.(9) it follows

$$\int_\Omega |\rho_n(x) - \rho(x)|^p \, dm(x) \leq b \tag{23}$$

$$\int_\Omega |\rho_n(x) - \rho(x)| \, dm(x) \leq \varepsilon_n \underset{n\to\infty}{\to} 0 \tag{24}$$

Here $b = b_1 + b_2$ for $0 < p < 1$ and $b = \left(b_1^{1/p} + b_2^{1/p}\right)^p$, see Eq.(7). By using Eqs.(23, 24) and the Lemma 5, with $a_1 = \varepsilon_n$, $a_p = b$, with $f(x) := \rho_n(x) - \rho(x)$, we get for $\min\ (1,p) < r < \max\ (1,p)$

$$\int_\Omega |\rho_n(x) - \rho(x)|^r \, dm(x) \leq \varepsilon_n^{\frac{p-r}{p-1}} b^{\frac{r-1}{p-1}} \tag{25}$$

$$[D_{r,m}[\rho_n - \rho]]^{\frac{1}{i(p)}} \leq \varepsilon_n^{\frac{p-r}{p-1}} b^{\frac{r-1}{p-1}} \tag{26}$$

From Eqs.(26, 7, 8, 10) we obtain

$$|D_{r,m}[\rho_n] - D_{r,m}[\rho]| \leq \delta_n \underset{n\to\infty}{\to} 0 \tag{27}$$

$$\delta_n = \left[\varepsilon_n^{\frac{p-r}{p-1}} m^{\frac{r-1}{p-1}}\right]^{i(p)} \tag{28}$$

Taking into account remark 2 we can summarize the previous results Eq.(27) as follows

Proposition 8 *Suppose that, for some $p \neq 1$, we have boundednes of $|\rho_n(x) - \rho(x)|$ in the L^p norm (Eq (23)) and convergence of the sequence $\rho_n(x)$ to $\rho(x)$ in L^1 norm Eq. (24). Then it follows the convergence in $L^r(\Omega, dm)$ distance for all values of r in the range* $\min\ (p,1) < r < \max\ (p,1)$

$$D_{r,m}[\rho_n(x) - \rho(x)] \underset{n\to\infty}{\to} 0 \tag{29}$$

In particular we have for all the Tsallis and Rényi entropies and functionals $D_{r,m}[]$ with r in this range

$$S_{T,r}[\rho_n] \underset{n\to\infty}{\to} S_{T,r}[\rho] \tag{30}$$

$$S_{R,r}[\rho_n] \underset{n\to\infty}{\to} S_{R,r}[\rho] \tag{31}$$

$$D_{r,m}[\rho_n(x)] \underset{n\to\infty}{\to} D_{r,m}[\rho(x)] \tag{32}$$

Remark 9 *From the previous result we note that the boundednes of the RE for some $p \neq 1$ has a stabilizing effect, only in the subset of all the PDFs satisfying Eq.(23). Hence, from the convergence in the natural L^1 distance we can conclude the convergence of the Rényi or Tsallis entropies. Note that from Eq.(25) and $p > 1$, when $r \nearrow p$ the error in the L^r norm increases and finally, when $r = p$ it attains the upper bound Eq.(23).*

From the point of view of statistical physics the convergence of sequence of PDF defined by some L^p distance is important because it follows the convergence of the expectation values of physical observable that belongs to the dual spaces [20]. Hence, in the case $p > 1$, from Eq.(29) we get the convergence $\|\rho_n - \rho\|_r \to 0$ for $1 < r < p$ and it follows the continuity of the expectation values of observable $f(x)$ from the dual space $f \in L^q(\Omega, dm)$ where $1/q + 1/r = 1$

$$\int_\Omega \rho_n(x) f(x) dm(x) \underset{n\to\infty}{\to} \int_\Omega \rho(x) f(x) dm(x) \tag{33}$$

$$f \in L^q(\Omega, dm);\ q > \frac{p}{p-1} \tag{34}$$

For $0 < q < 1$, it is possible that the dual space S of $L^q(\Omega, dm)$ is trivial. When S is not trivial (e.g., the space of sequences l^q) and $f \in S$), from L^p we obtain again the continuity of corresponding expectation values, similar to Eq.(33).

Continuity of RTE's when the PDF is approximated in L^p norm. Suppose that the *quality of approximation* $\rho_n(x)$ of the true PDF $\rho(x)$ is quantified in some L^p norm with $p \neq 1$. A convenient choice in the applications is $p = 2$. Suppose that $\rho(x),\ \rho_n(x) \in L^1(\Omega, m) \cap L^p(\Omega, m)$ and in analogy to the previous case we have

$$\int_\Omega |\rho_n(x) - \rho(x)|^p \, dm(x) \leq \varepsilon_n \underset{n\to\infty}{\to} 0 \tag{35}$$

$$\int_\Omega \rho_n(x) dm(x) = 1 \tag{36}$$

From the normalization condition, for ρ, ρ_n Eq.(2, 36) we obtain

$$\int_\Omega |\rho_n(x) - \rho(x)| \, dm(x) \leq 2 \tag{37}$$

By using the Lemma 5 with $a_1 = 2$ and $a_p = \varepsilon_n$ and Eq.(7) we obtain for all r in the domain

$$D = \{r|\min(1,p) < r < \max(1,p)\} \tag{38}$$

the following bounds

$$\int_\Omega |\rho_n(x) - \rho(x)|^r \, dm(x) \leq 2^{\frac{p-r}{p-1}} \varepsilon_n^{\frac{r-1}{p-1}} \underset{n\to\infty}{\to} 0 \tag{39}$$

$$[D_{r,m}[\rho_n - \rho]]^{\frac{1}{i(p)}} \leq 2^{\frac{p-r}{p-1}} \varepsilon_n^{\frac{r-1}{p-1}} \tag{40}$$

By using Eqs.(40, 7, respectively, 10) is

$$|D_{r,m}[\rho_n] - D_{r,m}[\rho]| \leq \delta_n \underset{n\to\infty}{\to} 0 \tag{41}$$

$$\delta_n := \left[2^{\frac{p-r}{p-1}} \varepsilon_n^{\frac{r-1}{p-1}}\right]^{i(p)} \tag{42}$$

From the Eqs.(41, 13-15) and Remark 2 we have the following stability results

Proposition 10 *Under the conditions Eqs.(2, 35, 36) for r in the range $\min(1,p) < r < \max(1,p)$ we have*

$$S_{T,r}[\rho_n] \underset{n\to\infty}{\to} S_{T,r}[\rho]$$
$$S_{R,r}[\rho_n] \underset{n\to\infty}{\to} S_{R,r}[\rho]$$

We recall that in the case when the measure $dm(x)$ is probabilistic, if $p > q \geq 1$ then $\|f\|_{L_q} < \|f\|_{L_p}$. If the measure $dm(x)$ is atomic, then from the convergence or boundednes in some l_p norm results the convergence or boundednes in all l_q norm, with $q > p$, including l_∞.

2.2.2 Boundednes and stability of the BSE

Details are given in Appendix 6.2

Counter example 1, discrete distribution We shall provide some counter examples that prove that the previous stabilizing conditions are not sufficient for the stability of the BSE. The details can be found in Appendix 6.2.1. Consider now the problem of continuity of the classical entropy, in the simplest case of the countable infinite probability space, with probabilities $\mathbf{p} := \{p_1, ..., p_n, ...\} := \{p_k\}_{k=1}^\infty$. In this case

$$S_{cl}[\mathbf{p}] = -\sum_{k=1}^{\infty} p_k \log p_k \tag{43}$$

Clearly

$$1 = \sum_{k=1}^{\infty} p_k = \|\mathbf{p}\|_{l^1} \tag{44}$$

and it is logical to consider that the *distance* between two probability laws $\mathbf{p} := \{p_1, ..., p_n, ...\}$ and $\mathbf{p}' := \{p_1', ..., p_n', ...\}$ is given by l^1 distance:

$$\|\mathbf{p} - \mathbf{p}'\|_{l^1} := \sum_{k=1}^{\infty} |p_k - p_k'|$$

Consider now the sequence that is convergent in the l^1

$$\mathbf{p}^{(n)} := \left\{ p_k^{(n)} \right\}_{k=1}^{\infty} \tag{45}$$

where

$$p_k^{(n)} := \frac{1}{K_n} \frac{1}{(k+4)\left[\log(k+4)\right]^{2+\frac{1}{n}}} \tag{46}$$

$$K_n = \sum_{k=1}^{\infty} \frac{1}{(k+4)\left[\log(k+4)\right]^{2+\frac{1}{n}}} < \infty \tag{47}$$

Note that $\left\|\mathbf{p}^{(n)}\right\|_{l^1} = 1$ and the limit of the sequence $\mathbf{p}^{(n)}$ in the space l^1 is the probability distribution given by $\mathbf{p} := \{p_k\}_{k=1}^{\infty}$ where

$$\mathbf{p} := \{p_k\}_{k=1}^{\infty} \tag{48}$$

$$p_k = \frac{1}{K} \frac{1}{(k+4)\left[\log(k+4)\right]^2} \tag{49}$$

$$K = \sum_{k=1}^{\infty} \frac{1}{(k+4)\left[\log(k+4)\right]^2} < \infty \tag{50}$$

We have also the following (general) inclusion:

$$\left\|\mathbf{p}^{(n)}\right\|_{l^1} = 1 \Rightarrow \mathbf{p}^{(n)} \in l^q \;; q > 1 \tag{51}$$

Despite the sequence $\mathbf{p}^{(n)}$ is convergent in l^1:

$$\left\|\mathbf{p}^{(n)} - \mathbf{p}\right\| \underset{n\to\infty}{\to} 0$$

it is easy to prove that the classical entropy is divergent:

$$S_{cl}[\mathbf{p}^{(n)}] = \mathcal{O}(n);\; n \to \infty \tag{52}$$

Consequently, the functional $\mathbf{p} \to S_{cl}[\mathbf{p}]$ from the space l^1 to $\mathbb{R}$ is not continuos, despite the classical entropy has the Lesche stability property and we have stabilizing condition from Eq.(51) in the form of l^p boundednes with $p > 1$.

Counter example 2, continuos distribution with finite measure m. A second counter example is the following (see also [32], page 7)

$$\rho_n(x) = \frac{M_n}{x\left(\log\frac{1}{x}\right)^{\alpha}};\ x \geq \frac{1}{n} \tag{53}$$

$$\rho_n(x) = 0;\ 0 < x < \frac{1}{n} \tag{54}$$

$$\rho(x) = \frac{M}{x\left(\log\frac{1}{x}\right)^{\alpha}} \tag{55}$$

$$1 < \alpha < 2 \tag{56}$$

where M_n, M are finite normalization constants. Note that (for further details see Appendix 6.2.2)

$$D_p(\rho_n) < C_p;\ D_p(\rho) < C_p \tag{57}$$

for some constants $C_p < \infty$, that does not depend on n. Consequently $\rho_n, \rho \in L^1 \cap L^p$ for all $0 < p \leq 1$ and

$$D_1[\rho_n - \rho] = \|\rho_n - \rho\|_{L^1} \underset{n\to\infty}{\to} 0 \tag{58}$$

so we have a wide choice of stabilizing conditions (see Proposition 8) assuring that $S_{R,q}(\rho_n) \to S_{R,q}(\rho)$ for all q in the range $p < q < 1$. Nevertheless, the classical entropy diverges:

$$S_{cl}[\rho_n] = -\mathcal{O}(\log n)^{2-\alpha} \to -\infty \tag{59}$$

2.2.3 Boundednes and stability of the BSE.

At the first stage, let us suppose in the following that the true PDF $\rho(x)$ is approximated by the sequence $\rho_n \in L^p(\Omega, dm) \cap L^q(\Omega, dm)$ where $0 < p < 1$ and $q > 1$ in the L^1 norm:

$$\int\limits_{\Omega} |\rho_n(x) - \rho(x)|\, dm(x) := \varepsilon_n \underset{n\to\infty}{\to} 0 \tag{60}$$

By Theorem 29 we have also $\rho_n \in L^1(\Omega, dm)$. Suppose, without loss of generality, that in addition we have the following stabilization conditions

$$Z_{p,m}[\rho_n] = D_{p,m}[\rho_n] \leq A;\ p < 1 \tag{61}$$

$$Z_{p,m}[\rho] = D_{p,m}[\rho] \leq A;\ p < 1 \tag{62}$$

$$Z_{q,m}[\rho_n] = D_{q,m}[\rho_n]^q \leq A;\ q > 1 \tag{63}$$

$$Z_{q,m}[\rho] = D_{q,m}[\rho]^q \leq A;\ q > 1 \tag{64}$$

For the sake of simplicity we will used a less strict bounds. These bounds are compatible with the normalization condition $Z_{1,m}[\rho_n] = 1$ if $1 \leq A$. From

Theorem 29, from the Appendix 6.4 and previous bounds, we get

$$Z_{r,m}[\rho_n] = [D_{r,m}[\rho_n]]^{1/i(r)} \leq A;\ p \leq r \leq q \tag{65}$$

$$Z_{r,m}[\rho] = [D_{r,m}[\rho]]^{1/i(r)} \leq A;\ p \leq r \leq q \tag{66}$$

From Eqs.(65, 66, 8) we extend these bounds to the complex domain, $p \leq \mathrm{Re}(w) \leq q$:

$$|Z_{w,m}[\rho_n]| = \left| \int\limits_{\Omega} |\rho_n(x)|^w \, dm(x) \right| \leq \int\limits_{\Omega} |\rho_n|^{\mathrm{Re}(w)} \, dm(x) \leq A \tag{67}$$

and in similar manner

$$|Z_{w,m}[\rho]| \leq A;\ p \leq \mathrm{Re}(w) \leq q \tag{68}$$

Boundednes In the following $Z_{q,m}[\rho]$ can be considered a particular value [48] of the analytic function $w \to Z_{w,m}[\rho]$, and related to the BSE by Eq.(16)

Proposition 11 *Under the conditions Eqs.[61,63] there exists an analytic continuation of the function $Z_{r,m}[\rho_n]$, denoted by*

$$F_n(z) := \int_{\Omega} |\rho_n(x)|^z \, dm(x) \tag{69}$$

in the strip from the complex z plane $D = \{z | p < \mathrm{Re}(z) < q\}$ such that $F_n(x) = Z_{x,m}[\rho_n]$ for $p \leq x \leq q$. We have for all $z \in D$ the bound

$$|F_n(z)| \leq A \tag{70}$$

The BSE of the PDF is given by the Cauchy integral

$$S_{cl}[\rho_n] = -\left[\frac{d}{dz} F_n(z) \right]_{z=1} = -\frac{1}{2\pi i} \oint\limits_{C} \frac{F_n(w) dw}{(w-1)^2} \tag{71}$$

where C is a sufficiently small circle centered in $w = 1$. The BSE is bounded by

$$|S_{cl}[\rho_n| \leq \frac{2A}{\min(q-1, 1-p)} \tag{72}$$

Proof. The Eq.(70) results from Corollary 30, Appendix (6.4). Eq.(71) is a direct consequence of Eqs (16, 69). To prove Eq.(72), we denote

$$r = \min(q-1, 1-p)/2 \tag{73}$$

the radius of the circle C, which is in the interior to the analyticity domain of $F_n(w)$. By Cauchy theorem and Eq.(71) we have

$$F_n(z) = \frac{1}{2\pi i} \oint\limits_{C} \frac{F_n(z')}{z' - z} dz' \tag{74}$$

$$S_{cl}[\rho_n] = -\frac{1}{2\pi i} \oint\limits_{C} \frac{F_n(z')}{(z'-1)^2} dz' \tag{75}$$

where C is the circle with radius r, so $|z'-1|=r$. By Eqs.(67, 68)

$$|F_n(z')| \leq A \tag{76}$$

From Eq.(75) we get

$$|S_{cl}[\rho_n]| \leq \frac{1}{r}\max_{z\in C'}|F_n(z)|$$

which combined with Eqs.(73, 76) completes the proof. ■

Stability of the BSE .

Consider now the stability problem. Suppose that the sequence of PDF ρ_n approximates the exact PDF ρ in the sense of Eq.(60) and we have the bounds Eqs.(61-64). We obtain the following result on the stability of the BSE:

Theorem 12 *Under the he previous conditions Eqs.(60-64) we have*

$$\lim_{n\to\infty} S_{cl}[\rho_n] = S_{cl}[\rho] \tag{77}$$

The proof (See Appendix 6.4.2) uses mathematical methods adopted in high energy physics [24]-[27].

In conclusion, the problem of numerical approximation of the classical and generalized entropies in the general case, cannot be solved without additional assumptions, that are tacitly used in practice. Among auxiliary assumptions that could stabilize the numerical instability in the computation of the entropies, we have smoothness conditions and bounds on the tail of the probability density function. We also mention that in many practical situation, the cumulative probability distribution function is approximated by the empirical cumulative distribution obtained from experiment. In this case the goodness of the fitting is characterized by a random variable, whose distribution is described by the Kolmogorov-Smirnov or the Anderson-Darling statistics. Consequently, the computed entropies are itself random variables and the problem of continuity must be formulated in terms of convergence of random variables. This class of problems deserves a further study.

3 The generalized Rényi entropies (GRE).

3.1 Motivations

Our generalization of the GRE defined in the previous work [10] is a straightforward extension of the previous case when the full phase space is a Cartesian product of two spaces, to the more general case of N factors. The line of reasoning is the same: first, we remark the mathematical relation between classical Rényi entropy and the metric in Lebesgue spaces and, second, we define the new entropies by using the metric in generalized Lebesgue spaces [30] (similar to the particular case $N=2$ studied in [10]). For characterization of generalized entropies from the point of view of fundamental mathematical structures, see [16].

3.1.1 Avoiding integrability problems

The PDF functions in many variables may have a more complex singularity structure such that all the RTE diverge. We proved in [10] that the GRE can solve this problem, in the case of PDF with two variables. The generalization introduced here is intended to treat similar integrability problems in many variables. As specified in the previous part, the domain of the entropy parameter q, where the Rényi or the Tsallis entropies are defined, is related to the singularities and to the asymptotic behavior of the PDF. Indeed, consider the case when the PDF depends on the real variable x so $(\Omega, \mathcal{A}, m) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda_1)$ where $\mathcal{B}(\mathbb{R})$ is the family of subsets of $\mathbb{R}$ generated by denumerable intersections and unions of open and close intervals, and λ_1 is the Lebesgue measure (more exactly $dm = d\lambda_1 = dx$). Suppose that the PDF $\rho(x)$ has the decomposition in a regular part, $\rho_0(x)$ is differentiable and decays, at least exponentially at infinity, in both singular part and a heavy tail part

$$\rho(x) = \rho_0(x)\left[1 + \frac{A}{|x|^{\alpha}}\right] + \frac{B}{1 + |x|^{1+\beta}}$$

where $A, B, \beta\ > 0$, and $0 < \alpha < 1$. The RTE is finite when $1/\left(1+\beta\right) < q < 1/\alpha$. In processing the experimental data, the large value of the RTE for some $q < 1$ is a signal characterized by a heavy tail of PDF (this is shown in the stationary PDF in self organized criticality models as well as in the linearized stochastic models with multiplicative noise [35]). The large value of RTE for some $q > 1$ suggests the existence of local singularity, that in some models is related to the phenomenon of noise driven intermittency [36]. Consequently, the plotting of the RTE is a good practice for detecting the existence of singularity and heavy tail effects.

Suppose now that we have a PDF depending of 3 variables $\rho(x_1, x_2, x_3)$ that similarly has a regular part ρ_0 that is differentiable and decay exponentially at infinity, and singular and heavy tail parts

$$\rho(x_1, x_2, x_3) = \rho_0(x_1, x_2, x_3)\left[1 + \sum_{i=1}^{3} \frac{A_i}{|x_i|^{\alpha_i}}\right] + \prod_{i=1}^{3} \frac{B_i}{1 + |x_i|^{1+\beta_i}} \tag{78}$$

with $\alpha_1 < \alpha_2 < \alpha_3$ and $\beta_1 < \beta_2 < \beta_3$. Note that the RTE associated to ρ from Eq.(78) is defined in the domain

$$1/\left(1+\beta_1\right) < q < 1/\alpha_3 \tag{79}$$

so it cannot detect the singularity (possible intermittency phenomena) in the variables x_1, x_2 and the heavy tail (possible SOC effects) in the variable x_2, x_3. In the sequel we will introduce a subsequent generalization that overcomes this difficulty.

3.1.2 Invariance under measure preserving transformations.

The classical as well as generalized entropies are general functionals that extracts/condense information about PDF, according to most general rules of the

probability theory. As a result the BSE, RTE are invariant under measure preserving transformations. For illustration we consider the simplest case, when we have a finite discrete probability distribution $p_{i,\alpha}$, with $1 \leq i \leq N$ and $1 \leq \alpha \leq A$, with $\sum\limits_{i=1}^{N}\sum\limits_{a=1}^{A} p_{i,\alpha} = 1$. Consider the case when the measure m in Eqs(8, 13) is the product of the counting measures on the sets $\{1,...,N\}$ and $\{1,...,A\}$. In the terminology of "Big Data", the NA data are presented in tensorized form [37]. The corresponding BSE and RE are

$$S_{cl} = -\sum_{i=1}^{N}\sum_{a=1}^{A} p_{i,\alpha}\log p_{i,\alpha} \tag{80}$$

$$S_{R,w} = \frac{1}{1-w}\log\sum_{i=1}^{N}\sum_{a=1}^{A} p_{i,\alpha}^{w};\ \ w > 0 \tag{81}$$

These entropies are invariant under the change of variables

$$p_{i,\alpha} \to p_{\sigma(i,\alpha)}$$

where the transformation $\sigma:\ \ (i,\alpha) \to \sigma(i,\alpha)$ is an element of the permutation group $\mathcal{S}_{NA}$ of NA objects, $\mathcal{S}_{NA}$ having $(NA)!$ elements, that reflect our complete lack of information about state space and ignore its Cartesian product structure. However, when the indices i and α have different physical meaning, such an extended symmetry hypothesis is not appropriate. On the other hand, the GRE's whose construction use the Cartesian product structure for generic $p_{i,\alpha}$ [10]

$$S_{v,w} := \frac{1}{1-w}\log\sum_{i=1}^{N}\left[\sum_{a=1}^{A} p_{i,\alpha}^{w}\right]^{v} \tag{82}$$

$$S_{p,q}^{(permuted)} := \frac{1}{1-w}\log\sum_{\alpha=1}^{A}\left[\sum_{i=1}^{N} p_{i,\alpha}^{w}\right]^{v} \tag{83}$$

is not invariant under full permutation group $\mathcal{S}_{NA}$. The knowledge of $S_{R,q}$ from Eq.(81) for all $w > 0$ allows to reconstruct the probabilities $p_{i,a}$ modulo permutation group $\mathcal{S}_{NA}$ (i.e. reconstruct the probabilities without specification of their place in the list). On the other hand the knowledge of the GRE's from Eqs.(82, 83) for all $v > 0$, $w > 0$ allows to reconstruct $p_{i,a}$ modulo smaller group $\mathcal{S}_N \times \mathcal{S}_A$ (see [16]). Similar problems appear in the case of probability distribution $p_{i,\alpha,m}$ where the indices i, α, m have different interpretation. The invariance group $\mathcal{S}_{NAM}$ of the Rényi entropy

$$S_{R,q} = \frac{1}{1-q}\log\sum_{i=1}^{N}\sum_{a=1}^{A}\sum_{m=1}^{M} p_{i,\alpha,m}^{q}$$

is too large, it contains $(NAM)!$ elements. In contrast, the generalized Rényi entropy, introduced in this work

$$\frac{1}{1-q_3}\log\sum_{i=1}^{N}\left[\sum_{a=1}^{A}\left[\sum_{m=1}^{M}p_{i,\alpha,m}^{q_3}\right]^{q_2}\right]^{q_1} \tag{84}$$

is no more invariant under full permutation group $\mathcal{S}_{NAM}$. On the other hand it is easy to see that the invariance group of Eq.(84) contains at least the product of subgroups, having at least $N!A!M!$ elements. For more detailed discussion, see [16]. In conclusion, while the RTE, BSE are constructed by using the most fundamental structures of the probability theory, the GRE also take into account the Cartesian product structure of the phase space and the corresponding product structures of measures.

3.2 Definitions and notations

We follow the same approach as in ref.[10]. We will define the Generalized Rényi entropies by using the results on Banach spaces with the anisotropic norm, exposed in ref.[30]. In the first part of the discussion we will restrict our discussions to the set of parameters that defines the GRE, when a) The integrals that appear in the definition can be interpreted as the distance in a suitable function space and b) The formula for entropy can be related to convexity or concavity properties of some functional, in the subspace of non negative density functions. Consequently, we will define only two class of distance-functionals and entropies, in analogy to functionals $S^{(1)}_{p_y,p_z}[\rho]$ and $S^{(2)}_{q_y,q_z}[\rho]$ defined in ref.[10].

Consider that the measure space $(\Omega, \mathcal{A}, m)$ has the following direct product structure. First, the phase space Ω is split into N subspaces

$$\Omega = \Omega_1 \times \Omega_2 \times ... \times \Omega_N \tag{85}$$

This means that the argument $\mathbf{x}$ of probability density function can be represented as $\mathbf{x} = \{x_1, x_2, ..., x_N\}$, so

$$\rho(\mathbf{x}) = \rho(x_1, x_2, ..., x_N) \tag{86}$$

with $x_k \in \Omega_k$, $1 \leq k \leq N$. We mention also that, in general, it is possible that the component spaces Ω_k are discrete, or has the structure of $\mathbf{R}^n$ or more in general, of an infinite dimensional measure space. Each of the spaces Ω_k has their $\sigma-$algebra $\mathcal{A}_k$. (The $\sigma-$algebra $\mathcal{A}$, that contains subsets of $\Omega = \Omega_1 \times \Omega_2 \times ... \times \Omega_N$ is defined as a tensor product of the $\sigma-$algebras $\mathcal{A}_k$: it is the smallest $\sigma-$algebra on Ω such that all of the projections $\Omega \xrightarrow{p_k} \Omega_k$ are measurable [19]-[21]). The measure m is also factorizable:

$$dm(\mathbf{x}) = dm(x_1, x_2, ..., x_N) = \prod_{j=1}^{N} dm_k(x_k) \tag{87}$$

where the measures m_k are defined on the $\sigma-$algebras (space of events) $\mathcal{A}_k$. In other words, the measure space $(\Omega, \mathcal{A}, m)$ is the direct product

$$(\Omega, \mathcal{A}, m) = \bigotimes_{j=1}^{N} (\Omega_j, \mathcal{A}_j, m_j) \tag{88}$$

The elementary probability $dP(\mathbf{x})$ is given by

$$dP(\mathbf{x}) = \rho(x_1, x_2, ..., x_N) dm(\mathbf{x}) \tag{89}$$

where $dm(\mathbf{x})$ is given by Eq.(87). Consider a vector $\mathbf{p} = \{p_1, p_2, ..., p_N\}$ of real numbers with $p_k \geq 1$. According to Ref.[30], in close analogy to Ref.[10] (where the particular case $N = 2$ was studied), we define recursively the anisotropic norm (depending on the measure m) $\|\rho\|_{\mathbf{p},m}$ as follows (see Appendix, subsection 6.3)

$$\rho_N(, x_1, x_2, ..., x_N) := \rho(x_1, x_2, ..., x_N) \tag{90}$$

$$\rho_{N-1}(x_{1,}, x_2, ..., x_{N-1}) := \left[\int\limits_{\Omega_N} [\rho_N \ (x_1, ..., x_N)]^{p_N} \ dm_N(x_N) \right]^{1/p_N} .... \tag{91}$$

$$\rho_{k-1}(x_1, x_2, ..., x_{k-1}) := \left[\int\limits_{\Omega_k} [\rho_k \ (x_1, ..., x_k)]^{p_k} \ dm_k(x_k) \right]^{1/p_k} ... \tag{92}$$

$$\rho_1 \ (x_1) := \left[\int\limits_{\Omega_2} [\rho_2 \ (x_1, x_2)]^{p_2} \ dm_2(x_2) \right]^{1/p_2} \tag{93}$$

$$\|\rho\|_{\mathbf{p},m} := \left[\int\limits_{\Omega_1} [\rho_1 \ (x_1)]^{p_1} \ dm_1(x_1) \right]^{1/p_1} \tag{94}$$

In analogy with Eqs.(4, 13) and Ref. [10] we define the GRE, with respect to the measure m

$$S_{\mathbf{p}}^{(1)}[\rho, m] = \frac{p_1}{1 - p_N} \log \|\rho\|_{\mathbf{p},m} \, ; p_i > 1 \tag{95}$$

Note that the also the anisotropic norm function $\rho \to \|\rho\|_{\mathbf{p},m}$ is convex and satisfies the axioms related to the norm [30]. The corresponding normed vector space is complete, i. e. it is a Banach space (see Ref.[30]). There is another range of parameters that generalizes the Rényi entropy corresponding to Eqs.(5, 14). Consider a vector $\mathbf{q} = \{q_1, q_2, ..., q_N\}$ of real numbers with $0 < q_k \leq 1$. In

analogy to Eqs.(90-94) we define recursively

$$\rho'_N(x_1, x_2, ..., x_N) := \rho(x_1, x_2, ..., x_N) \tag{96}$$

$$\rho'_{N-1}(x_1, x_2, ..., x_{N-1}) := \int\limits_{\Omega_N} \left[\rho'_M \ (x_1, ..., x_N)\right]^{q_N} dm_N(x_N).... \tag{97}$$

$$\rho'_{k-1}(x_1, x_2, ..., x_{k-1}) := \int\limits_{\Omega_k} \left[\rho'_k \ (x_1, ..., x_k)\right]^{q_k} dm_k(x_k)... \tag{98}$$

$$\rho'_1 \ (x_1) := \int\limits_{\Omega_2} \left[\rho'_2 \ (x_1, x_2)\right]^{q_2} dm_2(x_2) \tag{99}$$

$$N_{\mathbf{q},m}[\rho] := \int\limits_{\Omega_1} \left[\rho'_1 \ (x_1)\right]^{q_1} dm_1(x_1) \tag{100}$$

Observe that the mapping $\rho \to N_{\mathbf{q},m}[\rho]$ defines a pseudo-norm on the space of probability density functions (see Appendix, subsection 6.3). The map $\rho \to N_{\mathbf{q},m}[\rho]$ defines a concave function, in the subset of physically admissible PDF's, when $\rho(x_1, x_2, ..., x_N) \geq 0$. The GRE will be defined in analogy to Eqs.(5, 14) and to the case $N = 2$ from Ref. [10]

$$S^{(2)}_{\mathbf{q}}[\rho, m] = \frac{1}{1 - q_N} \log N_{\mathbf{q},m}[\rho]; \ 0 < q_i < 1 \tag{101}$$

For the sake of simplicity, we will use the extrapolated form of Eq. (101) in all range $q_k > 0$ of the parameters $\{q_1, ..., q_N\}$ allowing to relate $S^{(2)}_{\mathbf{q}}[\rho, m]$ to $S^{(1)}_{\mathbf{p}}[\rho, m]$. We obtain

$$S^{(2)}_{\mathbf{q}}[\rho, m] = S^{(1)}_{\mathbf{p}}[\rho, m] \tag{102}$$

$$N_{\mathbf{q},m}[\rho] = \left[\|\rho\|_{\mathbf{p},m}\right]^{p_1} \tag{103}$$

when p_i and q_i are related as follows

$$q_N = p_N \tag{104}$$

$$q_{N-1} = \frac{p_{N-1}}{p_N}... \tag{105}$$

$$q_k = \frac{p_k}{p_{k+1}}, ... \tag{106}$$

$$q_1 = \frac{p_1}{p_2} \tag{107}$$

Remark 13 *The algebraic equations for the Lagrange multipliers associated to maximal entropy problem are very complicated in the general case, nevertheless from the convexity or concavity properties the uniqueness of the solution follows. According to Eqs.(102-107), we are in the domain when* $\rho \to \|\rho\|_{\mathbf{p},m}$ *is a convex*

functional when

$$p_k = \prod_{j=k}^{N} q_j \geq 1 \tag{108}$$

In this case, the problem of maximal entropy with linear restriction is equivalent to minimization of a positive convex function and has unique solution. In the domain $0 < q_k < 1$*, where the map* $\rho \to N[\rho]_{\mathbf{q},m}$ *is a concave function, the generalized MaxEnt problem is equivalent to the maximization of a concave function with linear restriction. If the solution exists, it is unique. In the more general case exposed below the problem of uniqueness deserves further study.*

To have a more compact, more general extension of both the definitions of $\|\rho\|_{\mathbf{p},m}$ in Eqs.(90 -94) and $N_{\mathbf{q},m}[\rho]$ in Eqs.(96 -100), we define a more general distance functional for a more general range of parameters $\mathbf{p}$, that are generalizations of the functional defined in Eq.(7). We shall use the notation of Eq.(6), so the generalization of Eqs.(90-94, 96-100) are the following (for another equivalent construction see Appendix subsection 6.3)

$$f_N(, x_1, x_2, ..., x_N) := f(x_1, x_2, ..., x_N) \tag{109}$$

$$f_{N-1}(x_1,, x_2, ..., x_{N-1}) := \left[\int_{\Omega_N} |f_N(x_1, ..., x_N)|^{p_N} \, dm_N(x_N) \right]^{i(p_N)} \quad .... \tag{110}$$

$$f_{k-1}(x_1, x_2, ..., x_{k-1}) := \left[\int_{\Omega_k} |f_k \ (x_1, ..., x_k)|^{p_k} \, dm_k(x_k) \right]^{i(p_k)} \quad ... \tag{111}$$

$$f_1(x_1) := \left[\int_{\Omega_2} |f_2 \ (x_1, x_2)|^{p_2} \, dm_2(x_2) \right]^{i(p_2)} \tag{112}$$

$$D_{\mathbf{p},m}[f] := \left[\int_{\Omega_1} |f_1 \ (x_1)|^{p_1} \, dm_1(x_1) \right]^{i(p_1)} \tag{113}$$

We have the following important properties

Proposition 14 *The functional* $f \to D_{\mathbf{p},m}[f]$ *is a pseudo norm*

$$D_{\mathbf{p},m}[\alpha f] = |\alpha|^s \, D_{\mathbf{p},m}[f]\,; \ \alpha \in \mathbb{R} \tag{114}$$

$$D_{\mathbf{p},m}[f+g] \leq D_{\mathbf{p},m}[f] + D_{\mathbf{p},m}[g] \tag{115}$$

$$|D_{\mathbf{p},m}[f] - D_{\mathbf{p},m}[g]| \leq D_{\mathbf{p},m}[f-g] \tag{116}$$

where the homogeneity degree from Eq.(114) is $s = \prod_{k=1}^{N} [i(p_k)p_k]$

For proof see Appendix, Subsection 6.3

Always is it possible to relate the functionals $D_{\mathbf{p},m}[f]$ by $N_{\mathbf{q},m}[f]$. By comparing Eqs.(96-100) we obtain

$$D_{\mathbf{p},m}[f] = [N_{\mathbf{q},m}[f]]^{i(p_1)} \tag{117}$$

where the relation between exponents $\mathbf{q}=(q_1,...,q_N)$ and $\mathbf{p}=(p_1,...,p_N)$ is

$$q_N = p_N \tag{118}$$
$$q_{N-1} = p_{N-1}i(p_N) \tag{119}$$
$$q_{N-2} = p_{N-2}i(p_{N-1}) \tag{120}$$
$$... \tag{121}$$
$$q_2 = p_2 i(p_3) \tag{122}$$
$$q_1 = p_1 i(p_2) \tag{123}$$

From previous equations we have that the GRE can be expressed always either by functional $D_{\mathbf{p},m}[\rho]$ or by the functional $N_{\mathbf{q},m}[\rho]$

$$S^{(2)}_{\mathbf{q}}[\rho,m] = \frac{1}{1-q_N}\log N_{\mathbf{q},m}[\rho] = S^{(2)}_{\mathbf{q}}[\rho,m] = \frac{1}{i(p_{1)}(1-p_N)}\log D_{\mathbf{p},m}[f] \tag{124}$$

Remark 15 *Beside the previous definitions a more complete information about PDF can be obtained from* $S^{(2)}_{\mathbf{q}}[\rho^{(perm)},m]$ *where* $\rho^{(perm)}(x_1,...,x_N) := \rho(x_{T(1)},...,x_{T(N)})$ *and the map* $k\to T(k)$ *is a permutation of* N *indices.*

3.3 Properties of the GRE

3.3.1 Extensivity, in the classical sense

The extensivity follows from the multiplicative property of the norms $\|\rho\|_{\mathbf{p}}$ or pseudo-norms $N[\rho]_{\mathbf{q}}$. Suppose that for all of measure spaces $(\Omega_j,\mathcal{A}_j,m_j)$ that appear in Eq.(88) we have the splitting

$$(\Omega_j,\mathcal{A}_j,m_j) = (\Omega^{(1)}_j,\mathcal{A}^{(1)}_j,m^{(1)}_j)\otimes(\Omega^{(2)}_j,\mathcal{A}^{(2)}_j,m^{(2)}_j) \tag{125}$$

that means in particular that $\Omega_j = \Omega^{(1)}_j\times\Omega^{(2)}_j$, $x_j=\{x^{(1)}_j,x^{(2)}_j\}\in\Omega_j$,

$$dm_j(x_j) = dm_j(x^{(1)}_j,x^{(2)}_j) = dm^{(1)}_j(x^{(1)}_j)dm^{(2)}_j(x^{(2)}_j) \tag{126}$$

Accordingly we have the splitting of the phase space Ω

$$\Omega = \Omega^{(1)}\times\Omega^{(2)} \tag{127}$$
$$dm(\mathbf{x}) = dm^{(1)}(\mathbf{x}^{(1)})dm^{(2)}(\mathbf{x}^{(2)}) \tag{128}$$
$$\mathbf{x}^{(a)} = \{x^{(a)}_1,x^{(a)}_2...,x^{(a)}_N\};\ a=\overline{1,2} \tag{129}$$

where

$$\Omega^{(a)} = \Omega_1^{(a)} \times \Omega_2^{(a)} \times ... \times \Omega_N^{(a)}; \ a = \overline{1,2} \tag{130}$$

$$dm^{(a)}(\mathbf{x}^{(a)}) = \prod_{k=1}^{N} dm_j^{(a)}(x_j^{(a)}); \ a = \overline{1,2} \tag{131}$$

or in compact notation

$$(\Omega, \mathcal{A}, m) = (\Omega^{(1)}, \mathcal{A}^{(1)}, m^{(1)}) \otimes (\Omega^{(2)}, \mathcal{A}^{(2)}, m^{(2)}) \tag{132}$$

Suppose that the PDF is also factorized

$$\rho(\mathbf{x}) = \rho_1(\mathbf{x}^{(1)})\rho_2(\mathbf{x}^{(2)}) \tag{133}$$

Then we have the following relations, for all values of the parameters $\{p_1, ..., p_N\}$ or $\{q_1, ..., q_N\}$ such that the integrals make sense

$$\|\rho\|_{\mathbf{p},m} = \|\rho_1\|_{\mathbf{p},m_1} \|\rho_2\|_{\mathbf{p},m_2} \tag{134}$$

$$N_{\mathbf{q},m}[\rho] = N_{\mathbf{q},m_1}[\rho_1] N_{\mathbf{q},m_2}[\rho_2] \tag{135}$$

$$S_{\mathbf{p}}^{(1)}[\rho, m] = S_{\mathbf{p}}^{(1)}[\rho_1, m_1] + S_{\mathbf{p}}^{(1)}[\rho_2, m_2] \tag{136}$$

$$S_{\mathbf{q}}^{(2)}[\rho, m] = S_{\mathbf{q}}^{(2)}[\rho_1, m_1] + S_{\mathbf{q}}^{(2)}[\rho_2, m_2] \tag{137}$$

In the previous relations we defined $S_{\mathbf{p}}^{(1)}[\rho, m]$ and $S_{\mathbf{q}}^{(2)}[\rho, m]$ according to Eqs.(95, 101) and correspondingly

$$S_{\mathbf{p}}^{(1)}[\rho_a, m_a] = \frac{p_1}{1 - p_N} \log \|\rho_a\|_{\mathbf{p},m_a}; \quad a = \overline{1,2} \tag{138}$$

$$S_{\mathbf{q}}^{(2)}[\rho_a, m_a] = \frac{1}{1 - q_N} \log N_{\mathbf{q},m_a}[\rho_a]; \quad a = \overline{1,2} \tag{139}$$

For the sake of clarity, consider the following example with $N = 3$. Suppose $0 < q_1, q_2, q_3 \leq 1$ and $p_1, p_2, p_3 \geq 1$. We define $N_{\mathbf{q},m_a}[\rho_a]$ respectively $\|\rho_a\|_{\mathbf{p},m_a}$

with $a = \overline{1,2}$ as follows

$$N_{\mathbf{q},m_a}[\rho_a] = \int\limits_{\Omega_1^{(a)}} dm_1^{(a)}(x_1^{(a)}) \tag{140}$$

$$\left[\int\limits_{\Omega_2^{(a)}} dm_2^{(a)}(x_2^{(a)}) \left[\int\limits_{\Omega_3^{(a)}} dm_3^{(a)}(x_3^{(a)}) \rho \left(x_1^{(a)}, x_2^{(a)}, x_3^{(a)} \right)^{q_3} \right]^{q_2} \right]^{q_1} \tag{141}$$

$$\|\rho_a\|_{\mathbf{p},m_a}^{p_1} = \int\limits_{\Omega_1^{(a)}} dm_1^{(a)}(x_1^{(a)}) \tag{142}$$

$$\left[\int\limits_{\Omega_2^{(a)}} dm_2^{(a)}(x_2^{(a)}) \left[\int\limits_{\Omega_3^{(a)}} dm_3^{(a)}(x_3^{(a)}) \rho \left(x_1^{(a)}, x_2^{(a)}, x_3^{(a)} \right)^{p_3} \right]^{p_2/p_3} \right]^{p_1/p_2} \tag{143}$$

3.3.2 Particular cases.

In the following we will omit the measure, when no confusion arise: $\|\rho\|_{\mathbf{p},m} := \|\rho\|_{\mathbf{p}}$; $N[\rho]_{\mathbf{q},m} := N[\rho]_{\mathbf{q}}$; $S_{\mathbf{p}}^{(a)}[\rho, m] := S_{\mathbf{p}}^{(a)}[\rho]$. In the particular case when $p_1 = ... = p_N > 1$, or $0 < q_1 = q_2 = ... = q_{N-1} = 1$ and $q_N < 1$ the GRE is equal to the classical Rényi entropy from Eqs.(13, 14):

$$S_{\mathbf{p}}^{(1)}[\rho] = S_{R,p_N}[\rho] = \frac{p_N}{1-p_N} \log \|\rho\|_{p_N} \tag{144}$$

$$\|\rho\|_{p_N} = \left[\int\limits_{\Omega} dm(\mathbf{x}) \rho(\mathbf{x})^{p_N} \right]^{1/p_N} \tag{145}$$

respectively

$$S_{\mathbf{q}}^{(2)}[\rho] = S_{R,q_N}[\rho] = \frac{1}{1-q_N} \log N[\rho]_{q_N} \tag{146}$$

$$N_{q_N}[\rho] = \int\limits_{\Omega} dm(\mathbf{x}) \rho(\mathbf{x})^{q_N} \tag{147}$$

We used the notation from Eqs.(96-101). In particular when $p_N \searrow 1$ in Eqs.(144, 145) and when $q_1 = q_2 = ... = q_{N-1} = 1$; $q_N \nearrow 1$ in Eqs.(146, 147), respectively, we obtain the classical BSE

$$\lim_{p_1=...=p_N \searrow 1} S_{\mathbf{p}}^{(1)}[\rho] = \lim_{q_1=...=q_{N-1}=1;\ q_N \nearrow 1} S_{\mathbf{q}}^{(2)}[\rho] == -\int\limits_{\Omega} dm(\mathbf{x}) \rho(\mathbf{x}) \log \rho(\mathbf{x}) \tag{148}$$

3.3.3 Symmetry properties

Recall that the Rényi entropy is invariant under the group $\Gamma(\Omega, m)$ of invertible transformations of Ω that preserve the measure m. A measure preserving transformation $\Omega \overset{T}{\to} \Omega$ of the measure space $(\Omega, \mathcal{A}, m)$ is a transformation such that for all $A \subset \Omega$, $A \in \mathcal{A}$, we have $m\left[T^{-1}(A)\right] = m(A)$. Define the operator U_T acting on the distribution function as

$$\rho \to U_T \rho = \rho'$$
$$\rho'(x) := \rho\left[T(x)\right]$$

Then it is easy to verify that both Tsallis and Rényi entropies are invariant: for all ρ such that $S_{R,q}[\rho]$ is finite we have for all $T \in \Gamma(\Omega, m)$

$$S_{R,q}[\rho] = S_{R,q}[U_T \rho]$$

In the case of the classical definition of the Rényi entropy, when the measure space is discrete and the measure m is the counting measure, the group $\Gamma(\Omega, m)$ is the group generated by permutations of finite subsets of elements of Ω. Clearly, from Eq.(17), $S_{R,q}[\rho]$ is invariant under permutations. This property was one of the axioms in the axiomatic definitions of the classical Rényi entropy.
We denote by $\Gamma(\Omega_j\ , m_j)$ the group of measures preserving transformations of the measure space $(\Omega_j, \mathcal{A}_j, m_j)$ from the decomposition of $(\Omega, \mathcal{A}, m)$ from Eqs.(85 -88). Then we have the following

Proposition 16 *The GRE $S^{(1)}_{\mathbf{p}}[\rho]$, $S^{(2)}_{\mathbf{q}}[\rho]$, defined by Eqs.(95, 101), is invariant under the sub group $\Gamma(\Omega_1\ , m_1) \times \Gamma(\Omega_2\ , m_2) \times ... \times \Gamma(\Omega_N\ , m_N)$ of the full group $\Gamma(\Omega, m)$. Let T_k be a transformation of the space Ω_k that preserves the measure m_k. In other words $T_k \in \Gamma(\Omega_k\ , m_k)$. Define*

$$\rho'(x_1, ..., x_N) := \rho\left[T_1(x_1), ..., T_N(x_N)\right]$$

Then we have the invariance properties

$$S^{(1)}_{\mathbf{p}}[\rho] = S^{(1)}_{\mathbf{p}}[\rho']$$
$$S^{(2)}_{\mathbf{q}}[\rho] = S^{(2)}_{\mathbf{q}}[\rho']$$

with $S^{(1)}_{\mathbf{p}}[\rho]$, $S^{(2)}_{\mathbf{q}}[\rho]$ defined in Eqs.(95, 101).

Fore more details on the symmetry properties of GRE with respect to measure preserving transformations, see ref. [16], [13].

3.3.4 Geometric properties

Beyond the physical applications, the previous geometric definitions of the Rényi entropy and GRE are more advantageous; in our approach the basic objects are

the norms defined in Eqs.(90-94) or pseudo-norms defined in Eqs.(96 -100). In the case when $p_k \geq 1$, $\mathbf{p} = \{p_1, ..., p_N\}$, (see [30]) the norm $\|.\|_{\mathbf{p}}$ has the usual properties: for $a \in \mathbf{R}$ we have $\|a\rho\|_{\mathbf{p},m} = |a| \, \|\rho\|_{\mathbf{p},m}$, respectively [30]

$$\|\rho_1 + \rho_2\|_{\mathbf{p},m} \leq \|\rho_1\|_{\mathbf{p},m} + \|\rho_2\|_{\mathbf{p},m} \tag{149}$$

In particular, it follows the convexity of the mapping $\rho \to \|\rho\|_{\mathbf{p}}$: for $0 \leq \alpha \leq 1$ we have

$$\|\alpha\rho_1 + (1-\alpha)\,\rho_2\|_{\mathbf{p},\mathbf{m}} \leq \alpha\,\|\rho_1\|_{\mathbf{p},\mathbf{m}} + (1-\alpha)\,\|\rho_2\|_{\mathbf{p},m} \tag{150}$$

In the case $0 < q_k \leq 1$, $\mathbf{q} = \{q_1, ..., q_N\}$, the properties of the pseudo-norms $N[\rho]_{\mathbf{q},m}$ defined in Eqs.(96 -100) also allows geometrical interpretations. We have

$$N_{\mathbf{q},\mathbf{m}}[\rho_1 + \rho_2] \leq N[\rho_1]_{\mathbf{q},\mathbf{m}} + N[\rho_2]_{\mathbf{q},\mathbf{m}} \tag{151}$$

This can be proven recursively by using the definition and the simple inequality $|x+y|^q \leq |x|^q + |y|^q$, with $0 < q \leq 1$. Instead of convexity we have the following concavity inequality **in the first octant only** ($\rho_{1,2} \geq 0$)

$$N_{\mathbf{q},\mathbf{m}}[\alpha\rho_1 + (1-\alpha)\,\rho_2] \geq \alpha N_{\mathbf{q},\mathbf{m}}[\rho_1] + (1-\alpha)N_{\mathbf{q},\mathbf{m}}[\rho_2] \tag{152}$$

The Eq.(152) can be proven recursively by using the concavity of the function $f(x) := x^q$ with $0 < q \leq 1$.

By defining, the distance function between distribution functions ρ_1 and ρ_2 in the infinite dimensional space of PDF's by $d(\rho_1, \rho_2) := \|\rho_1 - \rho_2\|_{\mathbf{p}}$ for $p_k \geq 1$ and $d(\rho_1, \rho_2) := N_{\mathbf{q},\mathbf{m}}[\rho_1 - \rho_2]$ for $0 < q_k \leq 1$, respectively, we have the triangle inequality

$$d(\rho_1, \rho_3) \leq d(\rho_1, \rho_2) + d(\rho_2, \rho_3) \tag{153}$$

allowing the geometrical interpretation of GRE in term of distance in the functional space of admissible PDF's. The more general functional $D_{\mathbf{p},m}(\rho)$ is exposed in Appendix, subsection 6.3.

3.4 Applications of GRE.

In the work [10] was proven the H-Theorem: in the case $N = 2$, the GRE is a Liapunov functional for a class of random dynamical systems that describe the anomalous transport in plasmas [38], [39]. This property can be generalized easily for the $N > 2$ case. Recall that by using the Maximal Entropy principle for the RTE it is possible to obtain probability distribution functions with algebraic decay, similar to the derivation of the normal distribution [32]. By using the MaxEnt principle for GRE, a class of PDF with algebraic decay in one variable (for a particular parameter value it is the symmetric stable Cauchy-Lorentz distribution), and Gaussian in the second variable was derived [10]. From qualitative point of view such kind of mixed behavior is typical for the full PDF (the joint PDF of the driving and driven system) in random linear stochastic processes (see [35], [34]).

By computing the GRE of the full PDF of a complex dynamical system, that contains a Hamiltonian subsystem supposedly driven by an external dynamical

system, it is possible in a systematic way to detect the existence or absence of the back reaction [13].

In the case when $\Omega = \mathbb{R}^N$, $dm(x) = d^N x$, the N dimensional volume element, it is easy to prove that despite the Rényi entropy is sensitive to the rescaling, the variation of the Rényi entropy

$$S_{R,p,m}[\rho] - S_{R,q,m}[\rho]$$

is invariant with respect of the full group of affine transformation of the space $\mathbb{R}^N$, including rescaling, so it can used as a first criteria to identify distributions that differs by rescaling and Euclidean motions. This result can be generalized to GRE. Consider for instance in the case $N = 2$, $\Omega_k = \mathbb{R}^{N_k}$, $dm_k(x_k) = d^{N_k} x_k$ with $k = \overline{1,2}$. Let denote

$$\begin{aligned}
N_{q_1,q_2} &= \int\limits_{\mathbb{R}^{N_1}} d^{N_1}\mathbf{x}_1 \left[\int\limits_{\mathbb{R}^{N_2}} d\mathbf{x}_2 \rho(\mathbf{x}_1, \mathbf{x}_2)^{q_2} \right]^{q_1} \\
S_{q_1,q_2} &= \frac{1}{1-q_2} \log N_{q_1,q_2} \\
T_{q_1,q_2} &= S_{q_1,q_2} \frac{1-q_2}{q_1}
\end{aligned}$$

While $S_{q_1 q_2}$ is not invariant on the general affine group

$$\begin{aligned}
\mathbb{R}^{N_1} \ni \mathbf{x}_1 \to A_1\mathbf{x}_1 + \mathbf{b}_1 \in \mathbb{R}^{N_1} \\
\mathbb{R}^{N_2} \ni \mathbf{x}_2 \to A_2\mathbf{x}_2 + \mathbf{b}_2 \in \mathbb{R}^{N_2}
\end{aligned}$$

the linear combination of GRE's, with $\overline{q_1} \neq q_1$, $\overline{q_2} \neq q_2$, defined as follows

$$T_{\overline{q_1},\overline{q_2}} - T_{\overline{q_1},q_2} - T_{q_1,\overline{q_2}} + T_{q_1,q_2}$$

is an invariant and can be used to the classification of probability distributions.

A strategy to use the Rényi distribution is to compute for a large range of parameters. By Lemma on Rearrangements from [13], in the case of large class of measured spaces, from the equality

$$S_{R,q,m}[\rho_1] = S_{R,q,m}[\rho_2] \tag{154}$$

for all $q \in (a, b)$,results that ρ_1, ρ_2 are related by

$$\rho_1(x) = \rho_2(T(x)) \tag{155}$$

where $x \to T(x)$ is a map (not necessary continuos) that preserves the measure m. This observation allows to identify PDF's that are related by a measure-preserving coordinate change. Nevertheless there exists situations where the range of distributions ρ_2 that for given ρ_1 satisfy Eq.(154) is too large. Let a fixed measure space $(\Omega, \mathcal{A}, m)$ and consider two separate time series $x(t), y(t) \in$

Ω , that that is either a stochastic process or a deterministic process with random initial conditions, denote $x_k = x(t_k)$, $y_k = y(t_k)$. Denote by $\rho^{(x)}(x_1, x_2)$ respectively by $\rho^{(y)}(y_1, y_2)$ the corresponding joint PDF. If their Rényi entropies are equal, or, by Eq.(8, 13)

$$\int\limits_{\nleq} dx_1 \int\limits_{\nleq} dx_2 \rho^{(x)}(x_1, x_2)^q = \int\limits_{\nleq} dy_1 \int\limits_{\nleq} dy_2 \rho^{(y)}(y_1, y_2)^q$$

for all $q \in (a, b)$, according to Lemma on Rearrangements from [13] there exists a map $\Omega \times \Omega \ni (x_1, x_2) \to (y_1, y_2) = T(x_1, x_2) \in \Omega \times \Omega$ that preserves the measure $dm(x_1)dm(x_2) = dm(y_1)dm(y_2)$ such that

$$\rho^{(y)}(T(x_1, x_2)) = \rho^{(x)}(x_1, x_2) \tag{156}$$

The degree of indeterminacy from Eq.(156) can be reduced by if we compare the numerical values of the GRE corresponding to $N = 2$ case, or by Eqs.(96-101)

$$\int\limits_{\nleq} dx_1 \left[\int\limits_{\nleq} dx_2 \rho^{(x)}(x_1, x_2)^{q_2} \right]^{q_1} = \int\limits_{\nleq} dy_1 \left[\int\limits_{\nleq} dy_2 \rho^{(y)}(y_1, y_2)^{q_2} \right]^{q_1} \tag{157}$$

$$\int\limits_{\nleq} dx_2 \left[\int\limits_{\nleq} dx_1 \rho^{(x)}(x_1, x_2)^{q_2} \right]^{q_1} = \int\limits_{\nleq} dy_2 \left[\int\limits_{\nleq} dy_1 \rho^{(y)}(y_1, y_2)^{q_2} \right]^{q_1} \tag{158}$$

for $a_1 < q_1 < b_1$, $a_2 < q_2 < b_2$. According to Lemma on Rearrangements from [13] there exists a maps $\Omega_1 \ni x_1 \to y_1 = T_1(x_1) \in \Omega_1$ and $\Omega_2 \ni x_2 \to y_2 = T_2(x_2) \in \Omega_2$ that preserves the measures $dm(x_1) = dm(y_1)$, $dm(x_2) = dm(y_2)$ such that

$$\rho^{(y)}(T_1(x_1), T_2(x_2))) = \rho^{(x)}(x_1, x_2)$$

In conclusion, the use of GRE provide a finer classification of the PDF, classification that use measure theoretic aspects, in the case when we use the GRE for a large set of values of the parameters q_1, q_2.

In the case of analytic models of parametric destabilizations [33],[35], [34], gyrokinetic simulations of the micro instabilities in the tokamak plasma [47] or more generally, stochastic processes related to self-organized criticality [40]-[46], one of the problems is that at least in the theoretical models the mean value $\mathbb{E}(|x(t)|)$ is infinite or practically it is highly fluctuating. In this case the study of the long time correlation decay can be performed by studying, for example the GRE of the joint PDF for $x_k = x(t_k)$ where $k = \overline{1,3}$ with $t_2 - t_1 \to \infty$ and $t_3 - t_2 \to \infty$. In the case of stationary regime (see Eq.(140)) we have the

asymptotic factorization

$$N_{\mathbf{q}}[\rho] = \int_{\mathbb{R}} dx_1 \left[\int_{\mathbb{R}} dx_2 \left[\int_{\mathbb{R}} dx_3 \rho(x_1, x_2, x_3)^{q_3} \right]^{q_2} \right]^{q_1} \rightarrow$$
$$\left[\int_{\mathbb{R}} dx_2 \left[\rho_2(x_2)\right]^{q_2 q_3} \right]^{q_1} \left[\int_{\mathbb{R}} dx_3 \rho_3(x_3)^{q_3} \right]^{q_1 q_2} \int_{\mathbb{R}} dx_1 \left[\rho_1(x_1)\right]^{q_1 q_2 q_3}$$

where

$$\rho_1(x) = \int_{\mathbb{R}} dx_2 \int_{\mathbb{R}} dx_3 \rho(x, x_2, x_3), ...$$

or in term of entropies we have the decomposition at large time lags

$$\begin{aligned}(q_3 - 1) S^{(2)}_{q_1, q_2, q_3}[\rho] &\rightarrow q_1 q_2 (q_3 - 1) S_{R, q_3}[\rho_3] + \\ &q_1 (q_2 q_3 - 1) S_{R, q_2 q_3}[\rho_2] + (q_1 q_2 q_3 - 1) S_{R, q_1 q_2 q_3}[\rho_1]\end{aligned} \tag{159}$$

The speed of convergence in Eq.(159) can be used to characterize the correlation decay in stochastic processes, where the usual mean values diverge.

We remark also that in the cases when limiting values for $q_k \searrow 0$ and/or $q_j \rightarrow \infty$, the GRE give information on the support and extreme value properties of the PDF.

4 Bound and stability properties of GRE.

By the results and notations from subsection 6.4.3, the problem of boundednes of GRE can be treated directly. In the following we consider, according to the notations from subsection 3.2, the measure space $(\Omega, \mathcal{A}, m)$ Eq.(87, 88) and probability measure $dP(\mathbf{x}) = \rho dm$ from Eq.(89). Consider the set of exponents $(q_1, ..., q_N) := \mathbf{q}$ the associated functional $N_{\mathbf{q},m}$ from Eq.(100) and the hyper rectangle $D_N \subset \mathbb{R}^N$ defined in Eq.(300), Appendix 6.4.3 and suppose, for technical reasons, that we are in the generic case: D_N has non zero volume. Suppose in the continuation that at least one of the vertices of the hyper rectangle D_N contains the point $q_i = 1;\ 1 \leq i \leq N$, and denote this point by $\mathbf{u}$. Due to the normalization condition PDF we have $N_{\mathbf{u},m}(\rho) = 1$, so it is plausible to suppose that the functional $N_{\mathbf{q},m}$ is defined also in some neighborhood of $\mathbf{u} = (1, ...1)$. According to Theorem 31 the function $\mathbf{q} \rightarrow N_{\mathbf{q},m}(\rho)$ is log-convex in the variable $\mathbf{q}$. Suppose that on the vertices of the hyper rectangle D_N we have the bounds Eqs.(301), where in our case

$$g(w_1, ...w_N) = N_{\mathbf{w},m}(\rho) \tag{160}$$

Then, according to the Corollary 34, we have the bounds Eq.(309) (with the notations in subsection 6.4.3, Eqs.(302-307, 308)

$$\log N_{\mathbf{w},m}(\rho) \leq b_N(\mathbf{w});\ \mathbf{w} \in D_N \tag{161}$$

Denote by V_N the set of vertices of the hyper rectangle D_N and by V'_N the set of vertices excepting the vertex $\mathbf{u} = (1, ...1)$. Consider the exact PDF $\rho(\mathbf{x})$ and the approximating sequence $\rho_n(\mathbf{x})$

$$\int_{\Omega} dm(\mathbf{x}) \left|\rho_n(\mathbf{x}) - \rho(\mathbf{x})\right| dm(\mathbf{x}) \leq \varepsilon_n \underset{n\to\infty}{\to} 0 \tag{162}$$

or, equivalently, (Eqs. 96-100)

$$N_{\mathbf{u},m}(\rho_n - \rho) \leq \varepsilon_n \underset{n\to\infty}{\to} 0 \tag{163}$$

Suppose that on the rest of the vertices $\mathbf{v} \in V'_N$ we have the bounds

$$N_{\mathbf{v},m}(\rho_n) \leq B;\ \mathbf{v} \in V'_N \tag{164}$$

$$N_{\mathbf{v},m}(\rho) \leq B;;\ \mathbf{v} \in V'_N \tag{165}$$

where B is a constant. Denote by D'_N the subset of the hyper rectangle the set $Int\ (D_N)$, the set of interior points of D_N, defined by $Int(D_N) := \{\mathbf{w}|a_k^{(1)} < w_k < a_k^{(2)}\ ;\ k = \overline{1,N}\}$. By our previous technical assumption, D'_N is non void. The following stability result will be proved:

Proposition 17 *Under previous conditions Eqs.(162 -165) for all* $\mathbf{w} \in D'_N$ *we have*

$$N_{\mathbf{w},m}(\rho_n) \underset{n\to\infty}{\to} N_{\mathbf{w},m}(\rho) \tag{166}$$

$$S_{\mathbf{q}}^{(2)}[\rho_n, m] \underset{n\to\infty}{\to} S_{\mathbf{q}}^{(2)}[\rho, m] \tag{167}$$

Proof. The bounds from Eqs.(164, 165) can be translated in terms of distances $D_{\mathbf{p},m}[\rho]$, by using Eqs.(117 -123) :

$$D_{\mathbf{v},m}(\rho_n) \leq B';\ \mathbf{v} \in V'_N \tag{168}$$

$$D_{\mathbf{v},m}(\rho) \leq B';\ \mathbf{v} \in V'_N \tag{169}$$

$$B' = \max_{\mathbf{v}\in V'_N} \left(B^{i(v_1)}\right)$$

From Eqs.(115, 168, 169) results

$$D_{\mathbf{v},m}(\rho_n - \rho) \leq 2B';\ \mathbf{v} \in V'_N \tag{170}$$

This set of bounds we rewrite again in the term of $N_{\mathbf{w},m}$, by using Eqs.(117-123):

$$N_{\mathbf{v},m}(\rho_n - \rho) \leq B";\ \mathbf{v} \in V'_N \tag{171}$$

$$B" = \max_{\mathbf{v}\in V'_N} \left(2B'\right)^{1/i(v_1)}$$

Now we have bounds on all of the vertices of D_N and we can use Corrolary 34 and Eq.(308) with $g(\mathbf{w}) = N_{\mathbf{w},m}(\rho_n - \rho)$. We get:

$$\log N_{\mathbf{w},m}(\rho_n - \rho) \leq b_N(\mathbf{w});\ \mathbf{w} \in D'_N = Int(D_N) \tag{172}$$

$$b_N(\mathbf{w}) := \sum_{\mathbf{v}\in V_N} P(\mathbf{v},\mathbf{w}) \log A'_{\mathbf{v}} \tag{173}$$

$$0 < P(\mathbf{v},\mathbf{w}) < 1;\ \mathbf{w} \in D'_N \tag{174}$$

Because $V_N = V'_N \cup \{\mathbf{u}\}$ the Eq.(173) can be rewritten as follows

$$b_N(\mathbf{w}) = \sum_{\mathbf{v}\in V_N} P_{\mathbf{v}}(\mathbf{w}) \log\ A_{\mathbf{v}} \tag{175}$$

$$0 < P_{\mathbf{v}}(\mathbf{w}) < 1;\ \mathbf{w} \in D'_N \tag{176}$$

and $A_{\mathbf{v}}$ are the bounds on $g(\mathbf{w})$ on the vertices V_N. By using Eqs.(163, 171) we rewrite Eq.(175) as follows

$$\log N_{\mathbf{w},m}(\rho_n - \rho) \leq b_N(\mathbf{w}) = P_{\mathbf{u}}(\mathbf{w}) \log \varepsilon_n + K(\mathbf{w}); \mathbf{w} \in D'_N \tag{177}$$

$$K(\mathbf{w}) = \sum_{\mathbf{v}\in V'_N} P_{\mathbf{v}}(\mathbf{w}) \log\ B'' \tag{178}$$

From Eqs.(162, 176-178) results

$$N_{\mathbf{w},m}(\rho_n - \rho) \underset{n\to\infty}{\to} 0;\ \mathbf{w} \in D'_N$$

By using Eqs.(117, 123)

$$D_{\mathbf{w},m}(\rho_n - \rho) \underset{n\to\infty}{\to} 0;\ \mathbf{w} \in D'_N$$

From Eq.(116) we obtain

$$|D_{\mathbf{w},m}(\rho_n) - D_{\mathbf{w},m}(\rho)| \leq D_{\mathbf{w},m}(\rho_n - \rho) \underset{n\to\infty}{\to} 0;\ \mathbf{w} \in D'_N$$

and using again Eqs.(117, 123) we find the requested stability results

$$D_{\mathbf{w},m}(\rho_n) \underset{n\to\infty}{\to} D_{\mathbf{w},m}(\rho);\ \mathbf{w} \in D'_N$$
$$N_{\mathbf{w},m}(\rho_n) \underset{n\to\infty}{\to} N_{\mathbf{w},m}(\rho);\ \mathbf{w} \in D'_N$$

and from Eq.(124) and normalization of the PDF's ρ, ρ_n

$$S^{(2)}_{\mathbf{w}}[\rho_n, m] \underset{n\to\infty}{\to} S^{(2)}_{\mathbf{w}}[\rho_n, m];\ \mathbf{w} \in D'_N$$

which completes the proof. ■

Remark 18 *Note that it is possible the extend the proof to the case when the volume of D_N is zero, or to extend the stability proof to the part of boundary of D_N, where $P_{\mathbf{u}}(\mathbf{w}) > 0$.*

5 Conclusions

We proved that, in the general case of measure spaces, the entropies defined by C. Tsallis (TE), A. Rényi (RE) as well as the generalized Rényi's entropy (GRE), are well defined concepts and can be computed in a numerically stable manner for a large range of parameters, if a stabilizing condition is imposed. However, for the case of Shannon-Boltzmann's entropy (BSE) two stabilizing conditions are necessary. In all cases the stabilizing conditions are expressed in the term of finiteness of Lebesgue space L^p norm of the probability density functions. As a mathematical by-product, we proved the logarithmic convexity of the integrals related to generalized Lebesgue space norms.

Acknowledgement 19 *The authors are grateful to Prof. M. Van Schoor and Dr D. Van Eester from Royal Military School, Brussels. Gy. Steinbrecher is grateful to Prof. S. Ciulli, University Montpelier, G. Nenciu, Institute of Mathematics "Simion Stoilow" of the Romanian Academy, I. Sabba Ştefãnescu, University Karlsruhe and C. Pomponiu for useful discussions. Giorgio Sonnino is also grateful to Prof. P. Nardone and Dr. P. Peeters of the Université Libre de Bruxelles (ULB) for useful discussions and suggestions.*

6 Appendix

6.1 Analytic extrapolation theorem.

For easy reference, we shall prove in a special case the following theorem [24], [25], [26]. Denote by D the domain in the complex plain $\mathbb{C}$ defined as follows

$$D = \{z | z \in \mathbb{C},\ 0 < \mathrm{Im}(z) < b,\ a_1 < \mathrm{Re}(z) < a_2\} \tag{179}$$

The boundary of the domain is $\partial D = \Gamma_0 \cup \Gamma_2 \cup \Gamma_2 \cup \Gamma_3$ where

$$\begin{aligned}
\Gamma_0 &= \{z | z \in \mathbb{C},\ \mathrm{Im}(z) = 0,\ a_1 < \mathrm{Re}(z) < a_2\} \\
\Gamma_1 &= \{z | z \in \mathbb{C}, b \geq\ \mathrm{Im}(z) \geq 0,\ \mathrm{Re}(z) = a_1\} \\
\Gamma_2 &= \{z | z \in \mathbb{C},\ b \geq\ \mathrm{Im}(z) \geq 0,\ \mathrm{Re}(z) = a_2\} \\
\Gamma_3 &= \{z | z \in \mathbb{C},\ \mathrm{Im}(z) = b,\ a_1 < \mathrm{Re}(z) < a_2\}
\end{aligned}$$

Denote by $\mathcal{K}_\varepsilon$ the family of analytic functions in D, such that for all $f(z) \in \mathcal{K}_\varepsilon$ we have

$$|f(z)| \leq \varepsilon;\ z \in \Gamma_0 \tag{180}$$

$$|f(z)| \leq m;\ z \in \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \tag{181}$$

We shall prove the following theorem

Theorem 20 *Let $z_0 \in D$ and the constant m fixed. Under the previous conditions Eqs.(180, 181), for fixed m and for all $f(z) \in \mathcal{K}_\varepsilon$ we have*

$$|f(z_0)| \leq \delta(\varepsilon) \tag{182}$$

where $\delta(\varepsilon)$ does not depend on f and

$$\lim_{\varepsilon\to 0}\delta(\varepsilon) = 0 \tag{183}$$

Proof. Denote by $\Gamma_m := \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$ and by $u_0(z)$, $u_1(z)$ the harmonic functions in D with the following Dirichlet boundary conditions.

$$u_0(z) = 1;\ z \in \Gamma_0 \tag{184}$$

$$u_0(z) = 0;\ z \in \Gamma_m \tag{185}$$

$$u_1(z) = 1;\ z \in \Gamma_m \tag{186}$$

$$u_1(z) = 0;\ z \in \Gamma_0 \tag{187}$$

According to the standard terminology [49], [50], [27] the function $u_0(z)$ is the harmonic measure of the subset Γ_0 while $u_1(z)$ is the harmonic measure of Γ_m. We remark that

$$0 < u_k(z_0) < 1;\ k = 0, 1 \tag{188}$$

with strict inequalities, because z_0 is an interior point. This property is a consequence of the interpretation of $u_k(z_0)$ as hitting probability of planar Brownian motion [50]. These inequalities has also obvious physical meaning in the case when $u_k(z)$ are stationary temperature fields. Denote

$$U_\varepsilon(z) := u_0(z) \log \varepsilon + u_1(z) \log m \tag{189}$$

Also denote by $V_\varepsilon(z)$ the real harmonic conjugate of $U_\varepsilon(z)$, such that the function

$$K_\varepsilon(z) := U_\varepsilon(z) + iV_\varepsilon(z) \tag{190}$$

is an analytic function in D. Denote

$$C_\varepsilon(z) := \exp K_\varepsilon(z) \tag{191}$$

From Eqs.(184-191) we have that $C_\varepsilon(z)$ is analytic, without zeroes in D and have the following properties

$$|C_\varepsilon(z)| = \varepsilon;\ z \in \Gamma_0 \tag{192}$$

$$|C_\varepsilon(z)| = m;\ z \in \Gamma_m \tag{193}$$

$$\lim_{e\to 0} C_\varepsilon(z_0) = 0;\quad z_0 \in D \tag{194}$$

The function

$$G(z) := f(z)/C_\varepsilon(z) \tag{195}$$

is analytic in D and from Eqs.(180, 181, 192, 193) we have that $|G(z)| \leq 1$ for all $z \in \partial D$. From the maximum modulus principle [19] we get

$$|G(z)| \leq 1;\ \forall z \in \partial D \cup D \tag{196}$$

By denoting $\delta(\varepsilon) := C_\varepsilon(z_0)$, from Eqs.(195, 196) we have that $|f(z_0| \leq \delta(\varepsilon)$, which according to Eq.(194) completes the proof. ■

6.2 Some convergence results

6.2.1 Discrete case, counterexample 1, subsection 2.2.2

Proof of the inequalities (47, 50) It is sufficiently to prove Eq.(50), the convergence in Eq.(47) results by comparison: $K_n < K$. Because in Eq.(50) the sequence of p_k is monotone decreasing, we use the integral criteria: we have to prove that

$$K < \lim_{N\to\infty} \int_0^N dk \frac{1}{(k+4)\left[\log(k+4)\right]^2} < M \tag{197}$$

for some $M > 0$. By change of variable

$$k = \exp(x) - 4 \tag{198}$$

the Eq.(197) is reduced to the obvious bound for N large

$$\int_{\log(4)}^{\log(N+4)} \frac{dx}{x^2} < M$$

In conclusion $M = 1/\log(4)$ and

$$K_n < K < M \tag{199}$$

By considering only the first term in the infinite sum, we also obtain from Eq.(47),

$$\frac{1}{5[\log(5)]^{2+1/n}} < K_n \tag{200}$$

Proof of Eq.(52) We use Eqs.(43, 46, 47)

$$S_{cl}[\mathbf{p}^{(n)}] = A^{(n)} + B^{(n)} + C^{(n)} \tag{201}$$

$$A^{(n)} := \sum_{k=1}^{\infty} p_k^{(n)} \log K_n \tag{202}$$

$$B^{(n)} := (2 + \frac{1}{n}) \sum_{k=1}^{\infty} p_k^{(n)} \log\log(k+4) \tag{203}$$

$$C^{(n)} := \sum_{k=1}^{\infty} p_k^{(n)} \log(k+4) \tag{204}$$

From Eq.(44) we have $A^{(n)} = \log K_n$ and from Eq.(199, 200) we get

$$-\log[5(\log(5))^{2+1/n}] < A^{(n)} \leq \log M \tag{205}$$

We will denote by $C_{1,2,..}$ some fixed constants, that do not depend on n. From Eqs.(200, 203) we have

$$B^{(n)} < C_1 \sum_{k=1}^{\infty} \frac{1}{(k+4)\left[\log(k+4)\right]^2} \log\log(k+4) \tag{206}$$

Let N such that for all $k \geq N-1$ the summand in Eq.(206) is monotone decreasing. We obtain

$$B^{(n)} < C_2 + C_1 \sum_{k=N}^{\infty} \frac{1}{(k+4)\left[\log(k+4)\right]^2} \log\log(k+4) \tag{207}$$

In order to prove that the infinite sum in Eq.(207) is convergent we use the integral criteria and the change of variable Eq.(198)

$$B^{(n)} < C_2 + C_1 \int\limits_{N-1}^{\infty} dk \frac{1}{(k+4)\left[\log(k+4)\right]^2} \log\log(k+4) < \tag{208}$$

$$= C_2 + C_1 \int\limits_{\log(4)}^{\infty} dx \frac{\log(x)}{x^2} < +\infty \tag{209}$$

In conclusion we find that $B^{(n)}$ is uniformly bounded. The summand in Eq.(204) is monotone, so by the integral criteria we find

$$\int\limits_{0}^{\infty} dk \frac{1}{(k+4)\left[\log(k+4)\right]^{1+1/n}} < C^{(n)} < \int\limits_{1}^{\infty} dk \frac{1}{(k+4)\left[\log(k+4)\right]^{1+1/n}} \tag{210}$$

We use again Eq.(198) and we get

$$n \log(4) < C^{(n)} < n \log(5)$$

that together to Eqs.(201, 205, 209) prove Eq.(52)

6.2.2 Proof of the results from Counter Example 2, subsection 2.2.2

In the following we shall estimate, or explicitly compute, the integrals by using the substitution

$$x = \exp(-t) \tag{211}$$

The normalization constants M_n, M from Eqs(53-56) can be computed exactly by using Eq.(211)

$$M_n^{-1} = \frac{1}{\alpha - 1} \left[(\log\ 2)^{1-\alpha} - (\log\ n)^{1-\alpha} \right];\ n > 2 \tag{212}$$

$$M^{-1} = \frac{1}{\alpha - 1} (\log\ 2)^{1-\alpha} \tag{213}$$

It follows that $M_n \to M,\ \ \rho_n, \rho \in L^1,\ \|\rho_n\|_{L^1} = \|\rho\|_{L^1} = 1$ and by simple calculations and Eq.(211) results

$$\|\rho_n - \rho\|_{L^1} = \int\limits_0^{1/n} \frac{M dx}{x\left(\log \frac{1}{x}\right)^\alpha} + \int\limits_{1/n}^{1/2} \frac{|M_n - M| dx}{x\left(\log \frac{1}{x}\right)^\alpha} \to 0$$

In order to prove Eq.(57) it is sufficient to prove that for $0 < p < 1$

$$\int\limits_0^{1/2} \left[\frac{1}{x\left(\log \frac{1}{x}\right)^\alpha}\right]^p dx < \infty$$

resulting from the inequality $\log\ 1/x\ > \log\ 1/2$, for $0 < x < 1/2$, or by direct calculation by using Eq.(211).

In order to prove Eq.(59) we denote

$$f(x) := \frac{1}{x\left(\log \frac{1}{x}\right)^\alpha} \tag{214}$$

and from Eqs.(3, 53, 54, 214) we obtain

$$S_{cl}(\rho_n) = -\log M_n - M_n I_n \tag{215}$$

$$I_n = \int\limits_{1/n}^{1/2} f(x) \log f(x)\ dx \tag{216}$$

We use Eqs.(211, 214) so the last integral is rewritten as follows

$$I_n = \int\limits_{\log 2}^{\log n} \frac{dt}{t^\alpha} \left(t - \alpha \log t\right) \tag{217}$$

In the range $1 < \alpha < 2$ the integral $\int\limits_{\log 2}^{\infty} \frac{dt}{t^\alpha} \log t$ is convergent, so the leading term for $n \to \infty$ is given by

$$I_n \asymp \frac{1}{2-\alpha} \left[\log n\right]^{2-\alpha}$$

that proves Eq.(59)

6.3 Class of metric vector spaces and the functional $D_{\mathbf{p},m}[f]$, associated to GRE

We expose here a formalism in order to have a clear control on the pseudo-norms in the case of arbitrary number of variables. The notation from Eq.(6) will

be used in continuation. In analogy to the compact notation for the pseudo-norm defined in the case of single variables Eq.(7), we present here another, equivalent, definition with Eqs.(109-113) of the functional $D_{\mathbf{p},m}[f]$. We have in mind the measure space $(\Omega, \mathcal{A}, m)$ and a function $f(x_1, x_2, ..., x_N)$ similar to PDF $\rho(x_1, x_2, ..., x_N)$ with the structure specified in Eqs (85-89). We use the following

Definition 21 *Let E a Metric Vector Space (MVS) with a real-valued metric (distance to origin) $\delta : E \to \mathbb{R}_+$ such that for $x, y \in E$ the distance is $d(x,y) := \delta(x-y)$. The function $E \ni v \to \delta(v) \in \mathbb{R}_+$ is called pseudo-norm, if it satisfy the triangle inequality and it is homogenous with degree s*

$$\delta(u+v) \leq \delta(u) + \delta(v) \tag{218}$$

$$\delta(\alpha v) = |\alpha|^s \delta(v); \tag{219}$$

$$0 < s \leq 1 \tag{220}$$

$$\delta(v) = 0 \Rightarrow v = 0 \tag{221}$$

This metric vector space E with pseudo-norm δ will be denoted (E, δ).

Recall that in the case of norms in general we have $s = 1$ and in the case of pseudo norms $N_{\mathbf{p,m}}(\rho)$ from Eq.(100) $s = \prod_{k=1}^{N} q_k < 1$. In the case of distance defined previously in Eqs.(7, 113) $s = \prod_{k=1}^{N} ki(p_k) < 1$ (see below). Note that $s \leq 1$ results from Eqs.(218, 219). If $s < 0$, then $\delta(\frac{1}{n}v) = n^{-s}\delta(v) \to \infty$ for $n \to \infty$, and if $s = 0$ then $\delta(\frac{1}{n}v) = \delta(v)$. So in both cases the distance is discontinuous near $v = 0$.

Remark 22 *From Eq.(218) follows the useful inequality*

$$|\delta(u) - \delta(v)| \leq \delta(u-v) \tag{222}$$

We can define the convergence of a sequence of vectors u_n to u in E by the distance: $\delta(u_n - u) \to 0$. Then Eq.(222) means that the function $u \to \delta(u)$ is continuous: $u_n \to u \Rightarrow \delta(u_n) \to d(u)$.

6.3.1 The space $L^p(\Omega, m, E, d)$ and the functional $D_{p,m,E}(f)$.

In order to obtain an equivalent definition of $D_{\mathbf{p,m}}[f]$, from Eq.(113) we extend the definition of Eq.(7) in such a manner that the properties Eq. (6-11) of the pseudo-norm are preserved. We have the following

Definition 23 *Denote by $L^p(\Omega, m, F, \delta)$ the vector space of all functions defined on the measure space $(\Omega, \mathcal{A}, m)$ with values in the MVS (F, δ)*

$$f : \Omega \to F \tag{223}$$

such that

$$D_{p,m,F}(f) := \left[\int_{\Omega} \left[\delta\left[f\left(x\right)\right]\right]^{p} dm(x) \right]^{i(p)} < \infty \tag{224}$$

where δ is a pseudo-norm with homogeneity degree s.

We observe that in Eq.(224) can be obtained from Eq.(7) by performing the change

$$|f| \to \delta(f)$$

From Eqs.(218, 219 224) we have the following

Proposition 24 *The functional $D_{p,m,F}(f)$ from Eq.(224) is a pseudo-norm (see definition 21) and it is homogenous with degree $\sigma = spi(p)$ where s is the homogeneity degree of the pseudo-norm $\delta(.)$.*

Proof. The homogeneity results directly from Eqs.(219, 224). From Eqs.(218, 224) we get

$$D_{p,m,F}(f+g) \leq \left[\int_{\Omega} \left[\delta\left[f\right] + \delta\left[g\right]\right]^{p} dm(x) \right]^{i(p)} \tag{225}$$

In the case $0 < p < 1$, we have $i(p) = 1$ and by using the inequality $(|a|+|b|)^p \leq |a|^p + |b|^p$ we obtain

$$D_{p,m,F}(f+g) \leq D_{p,m,F}(f) + D_{p,m,F}(g) \tag{226}$$

In the case $p \geq 1$ Eq.(226) results from Minkowski inequality for L^p space norms ■

Note that according to Remark 22, we have

$$|D_{p,m,F}(f) - D_{p,m,F}(g)| \leq D_{p,m,F}(f-g) \tag{227}$$

6.3.2 Proof of the Proposition 14

The proof is by backward induction, in close analogy to the recurrent definition Eq.(109-113). At each step we define, according to Definition 23, a MVS (E_k, Δ_k). We shall denote by σ_k the homogeneity degree of the pseudo-norm Δ_k. In the first step we set in Definition 23 $\Omega = \Omega_N$, $m = m_N$, $p = p_N$, $F = \mathbb{R}$, $\delta(.) = |.|$. Define the corresponding MVS (E_N, Δ_N) as follows

$$E_N = L^{p_N}(\Omega_N, m_N, \mathbb{R}, |\ \ |) = L^{p_N}(\Omega_N, m_N) \tag{228}$$

$$\Delta_N(f_N) = \left[\int_{\Omega_N} |f_N|^{p_N} \, dm_N \right]^{i(p_N)} \quad ; \; f_N \in E_N \tag{229}$$

Note that, at this stage, (E_N, Δ_N) is the standard L^p MVS. We get the same conclusion by the Proposition 24: Δ_N is a pseudo-norm (for $p_N \geq 1$ it is the standard L^p norm), homogenous with degree $\sigma_N = p_N i(p_N)$.
In the second step, for the sake of clarity, we repeat the previous construction: we set in Definition 23 $\Omega = \Omega_{N-1}$, $m = m_{N-1}$, $p = p_{N-1}$, but we select, $F = E_N$, $\delta(.) = \Delta_N(.)$ from Eqs.(228, 229). Define the corresponding MVS (E_{N-1}, Δ_{N-1}) as follows.

$$E_{N-1} = L^{p_{N-1}}(\Omega_{N-1}, m_{N-1}, E_N, \Delta_N) \tag{230}$$

$$\Delta_{N-1}(f_{N-1}) = \left[\int_{\Omega_{N-1}} [\Delta_N(f_{N-1})]^{p_{N-1}} \, dm_{N-1} \right]^{i(p_{N-1})} \quad ; \; f_{N-1} \in E_{N-1} \tag{231}$$

By Proposition 24, Δ_{N-1} is a pseudo-norm with $\sigma_{N-1} = p_{N-1} i(p_{N-1}) \; \sigma_N$, so E_{N-1} is a MVS.

Now we proceed to the induction step. Consider that it is proven that Δ_k is a pseudo-norm with homogeneity degree σ_k and (E_k, Δ_k) is a MVS, with the structure:

$$E_k = L^{p_k}(\Omega_k, m_k, E_{k+1}, \Delta_{k+1}) \tag{232}$$

$$\Delta_k(f_k) = \left[\int_{\Omega_k} [\Delta_{k+1}(f_k)]^{p_k} \, dm_k \right]^{i(p_k)} \quad ; \; f_k \in E_k \tag{233}$$

according to the induction hypothesis. We construct again the MVS (E_{k-1}, Δ_{k-1}) : we set in Definition 23 $\Omega = \Omega_{k-1}$, $m = m_{k-1}$, $p = p_{k-1}$, and we select, $F = E_k$, $\delta(.) = \Delta_k(.)$ from Eqs.(232, 233). We obtain

$$E_{k-1} = L^{p_{k-1}}(\Omega_{k-1}, m_{k-1}, E_k, \Delta_k) \tag{234}$$

$$\Delta_{k-1}(f_{k-1}) = \left[\int_{\Omega_{k-1}} [\Delta_k(f_{k-1})]^{p_{k-1}} \, dm_{k-1} \right]^{i(p_{k-1})} \quad ; \; f_{k-1} \in E_{k-1} \tag{235}$$

By Proposition 24 Δ_{k-1} is a pseudo-norm with $\sigma_{k-1} = p_{k-1} i(p_{k-1}) \; \sigma_k$, so E_{k-1} is a MVS, that completes the induction step. The final MVS is $E_1 = L^{p_1}(\Omega_1, m_1, E_2, \Delta_2)$

By continuing this procedure down to $k = 1$ we obtain the pseudo-norm $\Delta_1(f_1) \equiv D_{\mathbf{p},m}[f_N]$ with $\sigma_1 = \prod_{k=1}^{N} p_k i(p_k)$ that completes the proof.

6.4 Logarithmic convexity related to the Rényi and generalized Rényi entropies

Typical examples of log-convex functions are the integrals that define the L^p norms as functions of the exponent p. Consequently, many of the properties of

the Tsallis and Rényi, as well as the generalized Rényi entropies can be derived from the properties of log-convex functions. We mention that in the case of single variable all of the results can be reduced to the well known facts [21]. Nevertheless we give here an self-contained treatment including also the case of single variable, because only our approach can be extended to the case of many variables. We have the following general

Definition 25 *A non-negative real valued function $g(\mathbf{w}) = g(w_1, ..., w_n)$, defined in the vector space $\mathbb{R}^n$ is log-convex if all of the one variable functions $w_k \to \log g(a_1, ..a_{k-1}, w_k, a_{k+1}, ..., a_n)$ are convex functions. Equivalently, every log-convex function can be represented as $\exp k(w_1, ..., w_n)$ where $k(w_1, ..., w_n)$ is a convex function* ***separately*** *in each of the variables, not necessary in the ensemble of all of the variables.*

We emphasize that the definition for many variables is adapted to our problem and it might differ from definitions in the standard textbooks.

6.4.1 Single variable

This part is used for the proof of results related to Rényi and Tsallis entropies, and it is the starting point for higher dimensional generalizations.

For a function of a single variable $g(w)$ that has continuos second derivative a necessary and sufficient condition for log-convexity is

$$\frac{d^2}{dw^2} \log g(w) = \frac{g(w)g''(w) - g'(w)^2}{g^2(w)} \geq 0 \tag{236}$$

The most important fact is the following well known interpolation property [21]. For easy reference, we give an elementary treatment

Proposition 26 *Let $w \to g(w)$ be a non-negative function of real variable w, defined at least on the interval $[a, b]$. If $g(w)$ is log-convex and*

$$g(a) \leq A; \;\; g(b) \leq B \tag{237}$$

then for all $w \in [a, b]$ we have the bound

$$g(w) \leq A^{\frac{b-w}{b-a}} B^{\frac{w-a}{b-a}} \tag{238}$$

Proof. The function $k(w) = \log f(w)$ is convex, so

$$k(\alpha a + (1-\alpha)b) \leq \alpha k(a) + (1-\alpha)k(b) \tag{239}$$

where $0 \leq \alpha \leq 1$. Take $\alpha = (b-w)/(b-a)$ and from Eq.(239) results

$$\log g(w) = k(w) \leq \frac{b-w}{b-a} k(a) + \frac{w-a}{b-a} k(b)$$

that with Eq.(237) results to be

$$\log g(w) = k(w) \leq \frac{b-w}{b-a}\log A + \frac{w-a}{b-a}\log B$$

that completes the proof. ■

We have the following

Proposition 27 *Consider in the measure space $(\Omega, \mathcal{A}, m)$ the family of functions $f(\mathbf{x}, w)$, such that for every fixed value of the variable $\mathbf{x} = \mathbf{x}_0 \in \Omega$ the one variable function $w \to f(\mathbf{x}_0, w)$ is log-convex (consequently $f(\mathbf{x}_0, w) > 0$), for each fixed value of the variable $w = w_0$, the function $\mathbf{x} \to f(\mathbf{x}, w_0)$ is integrable with respect to measure $dm(\mathbf{x})$ and the function $g(w)$*

$$g(w) := \int_\Omega dm(\mathbf{x}) f(\mathbf{x}, w) \tag{240}$$

has second derivative. Then $g(w)$ is log-convex. Equivalently, linear combination with non negative coefficients *of log-convex functions is log-convex too*

Proof. From the log-convexity of $f(x_0, w)$ results that for some function $k(x, w)$, which is convex in the variable w, we have

$$f(\mathbf{x}, w) = \exp k(\mathbf{x}, w) \tag{241}$$

According to Eq.(236) we have to prove

$$g(w)g''(w) - g'^2(w) \geq 0 \tag{242}$$

From Eqs.(240, 241, 242) results

$$g(w)\int_\Omega dm(\mathbf{x})\left\{\frac{\partial^2 k}{\partial w^2} + \left[\frac{\partial k}{\partial w}\right]^2\right\}\exp k(\mathbf{x}, w) - \left[\int_\Omega dm(\mathbf{x})\frac{\partial k}{\partial w}\exp k(\mathbf{x}, w)\right]^2 \geq 0 \tag{243}$$

The first term in Eq.(243) is always positive, since $k(\mathbf{x}, w)$ is convex in the variable w. For fixed w we define a new probability measure ($\int_\Omega dP(\mathbf{x}) = 1$) and introduce a new notation $h(\mathbf{x})$

$$dP(\mathbf{x}) := dm(\mathbf{x})\frac{\exp k(\mathbf{x}, w)}{g(w)} \tag{244}$$

$$h(\mathbf{x}) := \frac{\partial k(\mathbf{x}, w)}{\partial w} \tag{245}$$

By using the notations Eqs.(244, 245) and $\frac{\partial^2 k}{\partial w^2} \geq 0$, the proof is reduced to the well known inequality between mean value and mean square value of the random variable $h(\mathbf{x})$

$$\int_\Omega dP(\mathbf{x})\,[h(\mathbf{x})]^2 - \left[\int_\Omega dP(\mathbf{x})h(\mathbf{x})\right]^2 \geq 0 \tag{246}$$

that that proves Eq.(243). ■

Corollary 28 *Consider a measure space $(\Omega, \mathcal{A}, m)$ and the real valued function $\phi(y)$ defined on Ω, such that the function*

$$g(w) := \int_{\Omega} |\phi(x)|^w \, dm(x) \tag{247}$$

is finite on some interval $a \leq w \leq b$. Then $g(w)$ is log-convex.

Proof. It follows from the previous Proposition 27 with $f(x,w) = |\phi(x)|^w$ that is clearly log-convex. ■

From Corollary 28 and Proposition 26 it follows the following result. It can be deduced from known properties of the L^p norm in the case $1 \leq p < q$, or from the Hadamard three line theorem [21].

Theorem 29 *Suppose that $\phi(x) \in L^p(\Omega, m) \cap L^q(\Omega, m)$ with $0 < p < q$ and*

$$\int_{\Omega} |\phi(x)|^p \, dm(x) \leq A_p \tag{248}$$

$$\int_{\Omega} |\phi(x)|^q \, dm(x) \leq A_q \tag{249}$$

Then $\phi(x) \in L^r(\Omega, m)$ with $0 < p \leq r \leq q$ and

$$\int_{\Omega} |\phi(x)|^r \, dm(x) \leq A_q^{\frac{r-p}{q-p}} A_p^{\frac{q-r}{q-p}} \tag{250}$$

Corollary 30 *Under the condition of the previous theorem 29 there exists an unique analytic function $F(z) = \int_{\Omega} |\phi(x)|^z \, dm(x)$ defined in the strip $p < \mathrm{Re}(z) < q$ such that*

$$|F(x+iy)| \leq A_q^{\frac{x-p}{q-p}} A_p^{\frac{q-x}{q-p}}; \; p \leq x \leq q \tag{251}$$

The previous results are sufficient to have a simple proof of the interpolation Lemma5

Proof. In the previous Theorem 29 we put $\phi(x) = f(x)$, $p = \min(1, s)$ and $q = \max(1, s)$ ■

6.4.2 Proof of the stability of the BSE (Theorem 12)

In the first step of the proof, we shall obtain convergence bounds on the function $\int_{\Omega} |\rho_n(x)|^r \, dm(x) - \int_{\Omega} |\rho(x)|^r \, dm(x)$ on a subset of the interval $[p, q]$. From the bounds Eqs.(61-64) and triangle inequality Eq.(9), by simple algebra, we obtain the bounds

$$\int\limits_{\Omega} |\rho_n(x) - \rho(x)|^p \, dm(x) \leq B_p := 2A; \; p < 1 \tag{252}$$

$$\int\limits_{\Omega} |\rho_n(x) - \rho(x)|^q \, dm(x) \leq B_q := 2^q A; \; q > 1 \tag{253}$$

where B_p, B_q are constants. By using Eqs. (60, 252) and Theorem 29, with $\Phi(x) = \rho_n(x) - \rho(x)$ we obtain, for all r with $p \leq r \leq 1$

$$\int_\Omega |\rho_n(x) - \rho(x)|^r \, dm(x) \leq (\varepsilon_n)^{\frac{r-p}{1-p}} B_p^{\frac{1-r}{1-p}} \tag{254}$$

Similarly, from Eqs.(60, 253) and Theorem 29, for all r in the domain $1 \leq r \leq q$ the following inequality is obtained

$$\int_\Omega |\rho_n(x) - \rho(x)|^r \, dm(x) \leq B_q^{\frac{r-1}{q-1}} (\varepsilon_n)^{\frac{q-r}{q-1}} \tag{255}$$

In the following we will restrict ourselves to a smaller domain: let $p_1 = (1+p)/2$ and $q_1 = (1+q)/2$. Denote

$$d_{1,n} := \max_{r \in [p_1, 1]} (\varepsilon_n)^{\frac{r-p}{1-p}} B_p^{\frac{1-r}{1-p}} \tag{256}$$

$$d_{2,n} := \max_{r \in [1, q_1]} B_q^{\frac{r-1}{q-1}} (\varepsilon_n)^{\frac{q-r}{q-1}} \tag{257}$$

and remark that from Eqs.(255, 60), for all r, in the domain $p_1 \leq r \leq 1$, the following uniform convergence is found $(p_1 < 1 < q_1)$

$$\int_\Omega |\rho_n(x) - \rho(x)|^r \, dm(x) \leq d_{1,n} \underset{n\to\infty}{\rightarrow} 0 \tag{258}$$

Denote

$$G_n(r) := \int_\Omega |\rho_n(x)|^r \, dm(x) - \int_\Omega |\rho(x)|^r \, dm(x) \tag{259}$$

that, according to the Corollary 30, can be analytically continued to the strip $\{z | p \leq \operatorname{Re} z \leq q\}$. In the domain $p_1 \leq r \leq 1$, from the inequality (10) and Eq.(258) results for all $p_1 \leq r \leq 1$

$$|G_n(r)| \leq \int_\Omega |\rho_n(x) - \rho(x)|^r \, dm(x) \leq d_{1,n} \underset{n\to\infty}{\rightarrow} 0 \tag{260}$$

In a similar manner, in the domain $1 \leq r \leq q_1$, from the inequality(10) and Eq.(258) we have $\|\rho_n - \rho\|_r \leq [d_{2,n}]^{1/r}$ and from Eq.(10) we get

$$|\|\rho_n\|_r - \|\rho\|_r| \leq [d_{2,n}]^{1/r} \underset{n\to\infty}{\rightarrow} 0 \tag{261}$$

$$\|\rho_n\|_r = \left[\int_\Omega |\rho_n(x)|^r \, dm(x)\right]^{1/r} ; \; r \geq 1 \tag{262}$$

We have to obtain bounds on $G_n(r)$ on the interval $1 \leq r \leq q_1$. From Eq.(261) results that there exists N sufficiently large such that for all $n > N$ we have

$$A_r \leq \|\rho_n\|_r \leq B_r \tag{263}$$

where we denoted

$$A_r = \|\rho\|_r - d_{2,N}^{1/r} > 0 \tag{264}$$

$$B_r = \|\rho\|_r + d_{2,N}^{1/r} \tag{265}$$

Note that if $r \geq 1$, $A_r \leq x \leq B_r$, $A_r \leq y \leq B_r$ we have the following algebraic inequality

$$|x^r - y^r| \leq r B_r^{r-1} |x - y| \tag{266}$$

By setting $x = \|\rho\|_r$ and $y = \|\rho_n\|_r$ and by using Eqs.(261, 266) and the notation Eq.(259), we obtain, for all $1 \leq r \leq q_1$

$$|G_n(r)| \leq r B_r^{r-1} [d_{2,n}]^{1/r} \tag{267}$$

(Hint: use the mean value theorem for the function $x \to x^r$). Combined with Eq.(260), we obtain the following bound in the domain $p_1 \leq r \leq q_1$

$$|G_n(r)| \leq \delta_n \tag{268}$$

$$\delta_n = \max(d_{3,n}, d_{1,n}) \tag{269}$$

$$d_{3,n} = \max_{1 \leq r \leq q_1} r B_r^{r-1} [d_{2,n}]^{1/r} \tag{270}$$

From Eq.(260, 261, 267-270) the following uniform convergence bound results

$$|G_n(r)| \leq \delta_n \underset{n \to \infty}{\to} 0 \tag{271}$$

$$r \in \Gamma_0 := [p_1, q_1] \tag{272}$$

Let $b := q - p > 0$ and denote

$$\begin{aligned}
\Gamma_0 &:= \{z | p_1 \leq \operatorname{Re} z \leq q_1, \operatorname{Im} z = 0\} \\
\Gamma_1 &:= \{z | z = p_1 + it;\ 0 \leq t \leq b\} \\
\Gamma_2 &:= \{z | z = q_1 + it;\ 0 \leq t \leq b\} \\
\Gamma_3 &:= \{z | p_1 \leq \operatorname{Re} z \leq q_1, \operatorname{Im} z = b\} \\
\Gamma_m &:= \Gamma_1 \cup \Gamma_2 \cup \Gamma_3
\end{aligned}$$

and define the domain $D := \{z | p_1 < \operatorname{Re} z < q_1,\quad 0 < \operatorname{Im} z < b\}$ that is contained in the domain of holomorphy of G_n. Now we prepare to use the previous extrapolation Theorem 20. According to the Corollary 30 $G_n(r)$ can be analitically continued to the strip $\{z | p \leq \operatorname{Re} z \leq q\}$. From Eqs.(67, 68, 8) we get

$$|G_n(z)| \leq m = 2A;\ z \in D \cup \Gamma_0 \cup \Gamma_m \tag{273}$$

Similarly to the proof of the Proposition 11, consider now a circle C with the center at $z = 1$ and having radius $R = \min((1 - p_1)/2,\ (q_1 - 1)/2)$. From Eqs.(259, 71) we obtain

$$S_{cl}[\rho] = -\left[\frac{d}{dz} F(z)\right]_{z=1} = -\frac{1}{2\pi i} \oint_C \frac{F(w) dw}{(w-1)^2}$$

$$F(z) := \int_\Omega |\rho(x)|^z \, dm(x)$$

and finally results

$$S_{cl}[\rho_n] - S_{cl}[\rho] = -\frac{1}{2\pi i} \oint_C \frac{G_n(w)dw}{(w-1)^2} \tag{274}$$

It is clear that the convergence of the classical entropy is controlled by

$$|S_{cl}[\rho_n] - S_{cl}[\rho]| \leq \frac{1}{R} \max_{w \in C} |G_n(w)| \tag{275}$$

Now we use the Theorem 20, with the input data given in Eqs.(271, 272, 273). Then for all w in the half circle in the upper complex half plane we have

$$|G_n(w)| \leq K \exp[\log(\delta_n)\ u_0(w)];\ |w-1| = R;\ \mathrm{Im}(w) \geq 0 \tag{276}$$

where the harmonic function $u_0(w)$ was defined by Eqs.(184, 185), and K is a constant. Since $G_n(w)$ is real in the real axis, the previous Eq.(276) extends also in the lower semicircle. Since in the interior of the domain D the harmonic measure $u_0(w)$ is strictly positive, we have $G_n(w) \to 0$. From compactness of the circle, the uniform convergence $\max_{w \in C} |G_n(w)| \to 0$ is ensured, and finally we obtain that $|S_{cl}[\rho_n] - S_{cl}[\rho]| \underset{n\to\infty}{\to} 0$, which completes the proof.

6.4.3 Logarithmic convexity properties and bounds in the case of many variables.

Our definition of the log-convexity, in the case of many variables, differs from the usual definition: In the generalization to many variables we shall restrict our study to the case when a function of N variables is log-convex in one of variables while the remaining $N-1$ are fixed. Typical example of interest is the nested integral that appear in the definition of the GRE. We will study the function of many variables $\mathbf{w} = (w_{1,} w_2, ..., w_N)$, that is defined recurrently in Eqs.(96-100). We set

$$g(w_{1,} w_2, ..., w_N) := N_{\mathbf{w},m}[\rho] \tag{277}$$

By using the previous Corollary 28 it is clear that for fixed $w_2^{(0)}, ..., w_N^{(0)}$, the function $w \to g(w, w_2^{(0)}, ..., w_N^{(0)})$ is log-convex. We have the following more general result

Theorem 31 *For all $1 \leq k \leq N$, for fixed $w_1^{(0)} = q_1, ..., w_{k-1}^{(0)} = q_{k-1}, w_{k+1}^{(0)} = q_{k+1}, ..., w_N^{(0)} = q_N$, the function $w \to g(w_1^{(0)}, ..., w_{k-1}^{(0)}, w, w_{k+1}^{(0)}, ..., w_N^{(0)}) := h(w)$ is log-convex in the variable w.*

Proof. The proof is by induction. We rewrite the part of interest from the reccurence relations Eqs.(96-100) as follows

$$\rho'_k(x_1, x_2, ..., x_k) := \int\limits_{\Omega_k} \left[\rho'_{k+1}\,(x_1, ..., x_{k+1})\right]^{q_{k+1}} dm_{k+1}(x_{k+1}) \tag{278}$$

$$\rho'_{k-1}(w, x_1, x_2, ..., x_{k-1}) := \int\limits_{\Omega_k} \left[\rho'_k\,(x_1, ..., x_k)\right]^{w} dm_k(x_k) \tag{279}$$

$$\rho'_{k-2}(w, x_1, x_2, ..., x_{k-2}) := \int\limits_{\Omega_k} \left[\rho'_k\,(x_1, ..., x_k)\right]^{q_{k-1}} dm_k(x_k) \tag{280}$$

$$... \tag{281}$$

$$\rho'_1\,(w, \mathbf{x}_1) := \int\limits_{\Omega_2} \left[\rho'_2\,(w, x_1, x_2)\right]^{q_2} dm_2(x_2) \tag{282}$$

$$h(w) := \int\limits_{\Omega_1} \left[\rho'_1\,(w, x_1)\right]^{q_1} dm_1(x_1) \tag{283}$$

From Corollary 28 results that the function $\rho'_{k-1}(w, x_1, x_2, ..., x_{k-1})$, given by Eq.(279), is log-convex in the variable w. By successive application of the Proposition 27 to Eqs.(280-283) and by observing that any positive power of a log convex function is log-convex, we conclude by induction, successively, that the functions $\rho'_{k-2}(w, x_1, x_2, ..., x_{k-2})$,..., $\rho'_1\,(w, \mathbf{x}_1)$, $h(w)$ are all log-convex in the variable w. ■

For the sake of clarity we consider first the case $N = 2$. Suppose now that $g(w_1, w_2)$ is log-convex in the variables w_1, w_2, in the rectangular domain D_2

$$a_1^{(1)} \leq w_1 \leq a_1^{(2)}; \quad a_2^{(1)} \leq w_2 \leq a_2^{(2)} \tag{284}$$

and we have the following bounds in the corner points of D_2

$$g\left(a_1^{(1)}, a_2^{(1)}\right) \leq A_{1,1} \tag{285}$$

$$g\left(a_1^{(2)}, a_2^{(1)}\right) \leq A_{2,1} \tag{286}$$

$$g\left(a_1^{(1)}, a_2^{(2)}\right) \leq A_{1,2} \tag{287}$$

$$g\left(a_1^{(2)}, a_2^{(2)}\right) \leq A_{2,2} \tag{288}$$

We denote

$$\rho_1^{(1)}(w_1) = \frac{a_1^{(2)} - w_1}{a_1^{(2)} - a_1^{(1)}} \tag{289}$$

$$\rho_1^{(2)}(w_1) = \frac{w_1 - a_1^{(1)}}{a_1^{(2)} - a_1^{(1)}} \tag{290}$$

$$\rho_2^{(1)}(w_2) = \frac{a_2^{(2)} - w_2}{a_2^{(2)} - a_2^{(1)}} \tag{291}$$

$$\rho_2^{(2)}(w_2) = \frac{w_2 - a_2^{(1)}}{a_2^{(2)} - a_2^{(1)}} \tag{292}$$

Remark 32 *If* $\mathbf{w} = \{w_1, w_2\}$ *is an interior point of the rectangle* D_2 *then* $0 < \rho^{(i)}(w_k) < 1$*, for all* $1 \le i \le 2$ *and* $1 \le k \le 2$

By using successively the Proposition 26, we define the following bound in the interior point with coordinates (w_1, w_2), in terms of values on the corner points of the rectangle Eq.(284):

$$\log\ g(w_1, w_2) \le b_2(w_1, w_2) := b_2(\mathbf{w}) \tag{293}$$

$$b_2(\mathbf{w}) = P_{1,1}(\mathbf{w}) \log A_{1,1} + \ P_{1,2}(\mathbf{w}) \log A_{1,2} \tag{294}$$

$$+ P_{2,1}(\mathbf{w}) \log A_{2,1} + P_{2,1}(\mathbf{w}) \log A_{2,2} \tag{295}$$

$$P_{m_1,m_2}(\mathbf{w}) = P_{m_1,m_2}(w_1, w_2) {=} \rho_1^{(m_1)}(w_1)\rho_2^{(m_2)}(w_2); \quad m_1, m_2 = \overline{1,2} \tag{296}$$

From the previous remark results $0 < P_{m_1,m_2}(\mathbf{w}) < 1$ if $\mathbf{w}$ is an interior point of the rectangle D_2

From Eq.(293) we obtain the following generalization of Proposition 26

Proposition 33 *Suppose that we have the sequence of* n *functions* $g_n(w_1, w_2)$ *that are log convex (hence non negative) in the domain Eq.(284) and on the* 4 *corner points we have the uniform bound*

$$g_n(a_1^{(i)}, a_2^{(j)}) \le A_{i,j};\ \ i, j = \overline{1,2} \tag{297}$$

Then in the all interior points of the rectangle Eq.(284) we have the uniform bound

$$g_n(w_1, w_2) \le \exp b(w_1, w_2)$$

with $b(w_1, w_2)$ *given by Eq.(294). If in at least in one of the corner points* $(a_1^{(i_0)}, a_2^{(j_0)})$ *we have*

$$g_n(a_1^{(i_0)}, a_2^{(j_0)}) \to 0 \tag{298}$$

and in the rest of the corner points we have the uniform bound Eq.(297) then for all interior points (w_1, w_2) *we have also*

$$g_n(w_1, w_2) \to 0 \tag{299}$$

for all interior *points in the rectangle defined by Eq.(284).*

Now we extend the previous result for arbitrary number of variables. The relations that follows in the particular case $N = 2$ reduces to Eqs.(284-299). Suppose that $g(w_1, w_2, ..., w_N)$ is log-convex in the variables $w_1, w_2, ..., w_N$, in the N dimensional hyper rectangle domain $D_N \subset \mathbb{R}^N$

$$a_k^{(1)} \leq w_k \leq a_k^{(2)} \quad ; \ k = \overline{1, N} \tag{300}$$

Suppose that in the 2^N corner points of D_N (the set of vertices V_N of the hyper-rectangle) we have the bounds

$$g(a_1^{(v_1)}, a_2^{(v_2)}, ..., a_N^{(v_N)}) \leq A_{v_1, v_2, ..., v_N}; \quad v_j = \overline{1, 2}; \quad j = \overline{1, N} \tag{301}$$

or denoting $v := (v_1, ..., v_N) \in \{1, 2\}^N$, the set of vertices V_N are represented as follows: $(a_1^{(v_1)}, a_2^{(v_2)}, ..., a_N^{(v_N)}) := \mathbf{a}^{(\mathbf{v})} \in V_N$. In analogy with Eqs.(289-292) we introduce the following notations

$$\rho_k^{(1)}(w_k) = \frac{a_k^{(2)} - w_k}{a_k^{(2)} - a_k^{(1)}} \quad ; \ k = \overline{1, N} \tag{302}$$

$$\rho_k^{(2)}(w_k) = \frac{w_k - a_k^{(1)}}{a_k^{(2)} - a_k^{(1)}} \quad ; \ k = \overline{1, N} \tag{303}$$

To obtain a more compact notation for the generalization of Eqs.(293-296), we use the following new notations for the set of constants $A_{v_1, v_2, ..., v_N}$

$$A'_{\mathbf{v}} := A_{v_1, v_2, ..., v_N} \quad v_j = \overline{1, 2}; \quad j = \overline{1, N} \tag{304}$$

and define the function $P : V_N \times D_N \to \mathbb{R}$ as follows

$$P(\mathbf{v}, \mathbf{w}) := \prod_{k=1}^{N} \rho_k^{(v_k)}(w_k); \quad v_j = \overline{1, 2}; \quad j = \overline{1, N} \tag{305}$$

$$v := (v_1, ..., v_N) \in \{1, 2\}^N \tag{306}$$

$$\mathbf{w} = (w_1, ..., w_N) \tag{307}$$

By Remark 32 we have

$$0 < P(\mathbf{v}, \mathbf{w}) < 1; \ \mathbf{w} \in Int(D_N); \ \mathbf{v} \in \mathbf{V}_N \tag{308}$$

where the set of interior points $Int(D_N) \subset D_N$ is defined by $a_k^{(1)} < w_k < a_k^{(2)} \ ; \ k = \overline{1, N}$. By using successively the Proposition 26, with the notations Eqs.(304-307) the following bounds results

Corollary 34 *If the function $\mathbf{w} \to g(\mathbf{w})$ is log-convex, then under previous conditions Eqs.(301-303) we have the bound in the hyper-rectangle D_N*

$$\log \ g(w_1, w_2, ..., w_N) \leq b_N(\mathbf{w}) \tag{309}$$

where we denote

$$b_N(\mathbf{w}) := \sum_{\mathbf{v} \in V_N} P(\mathbf{v}, \mathbf{w}) \log A'_{\mathbf{v}} \tag{310}$$

Remark 35 *It can be proven that* $\sum_{\mathbf{v}\in V_N} P(\mathbf{v},\mathbf{w}) = 1$

References

[1] C. E. Shannon, *A Mathematical Theory of Communication*, *Bell System Technical Journal*, **27**, 379-423 & 623-656 (1948).

[2] A. Rényi (1960), *On measures of information and entropy*, *Proceedings of the fourth Berkeley Symposium on Mathematics, Statistics and Probability, June 20-July 30, 1960, Volume I, University of California Press, Berkeley and Los Angeles, 547-561 1960*. pp. 547–561.

[3] A. Rényi, Wahrscheinlichkeitstheorie (De Gruyter, Berlin, 1974).

[4] A. Rényi, *On statistical laws of accumulation of information*, in *Rényi Alfréd Válogatott Munkái (A. Rényi Selected works)*, volume 3, 1962-1970, Akadémiai Kiadó, Budapest 1977, page 33.

[5] A. Rényi, Rev. Int. Inst. Stat., **33**, 1-14 (1965).

[6] C. Tsallis, *Journal of Statistical Physics*, **52**, 479–487 (1988).

[7] C. Tsallis, R. S. Mendes and A. R. Plastino, Physica **A, 261**, 534-554 (1998).

[8] C. Tsallis, *Introduction to nonextensive statistical mechanics_ approaching a complex world, (*Springer, 2009).

[9] M. Gell-Mann, C. Tsallis, *Nonextensive entropy- Interdisciplinary applications,* (Oxford University Press, USA, 2004).

[10] G. Sonnino, G. Steinbrecher, Phys. Rev. **E 89**, 062106 (2014).

[11] Yu. L. Klimontovich, Chaos, Solitons, Fractals, **5**, 1985-2002 (1995).

[12] I. Csiszár, I.E.E.E. Transactions on Information Theory, **41** -1, 26, (1995).

[13] G. Steinbrecher, G. Sonnino, "*Generalized Rényi Entropy and Structure Detection of Complex Dynamical Systems*", arXiv:1512.06108v1 [physics.data-an] (2015).

[14] T. Maszczyk and X. Duch, *Comparison of Shannon, Rényi and Tsallis Entropy used in Decision Trees*, *Lecture Notes in Computer Science*, **5097**, 643 (2008).

[15] T. van Erven, P. Harremoës, "Rényi Divergence and Kullback-Leibler Divergence", e-print arXive:1206.2459v2 (2014).

[16] G. Steinbrecher, A. Sonnino, G. Sonnino, Journal of Modern Physics, **7**, 251-266 (2016).

[17] Hongfei Cui, Jianqiang Sun, Yiming Ding, *The rates of convergence for generalized entropy of the normalized sums of IID random variables*, arXiv:1106.3381v1 [cs.IT] (2011).

[18] B. Lesche, Journal of Statistical Physics, **27**, 419-422, (1982).

[19] W. Rudin (1987), Real and Complex Analysis, McGraw Hill Inc. 3rd Ed. page 74.

[20] M. Reed and B. Simon, Functional Analysis (Methods of Modern Mathematical Physics). Vol. 1, Academic Press (1981).

[21] E. M. Stein, R. Shakarchi, *Functional Analysis, Introduction to further topics, Princeton Lectures in Analysis IV, Princeton Univ. Press, Princeton (2011),page 36 exerc. 9.c. page 39 exerc. 20.*

[22] M. Ciulli, S. Ciulli, Computer Physics Communications, **18**, 215 (1981).

[23] I. Caprini, M. Săraru, C. Pomponiu, M. Ciulli, S. Ciulli, I. Sabba-Ştefãnescu, Computer Physics Communications **18**, 305-326 (1979).

[24] S. Ciulli, *Stability Problems in Analytic Continuation*, Lectures given at the 1972 Int. Institute for theoretical strong interaction physics, Kaiserslautern. Pag. 70-105, Springer Berlin Heidelberg (1973).

[25] S. Ciulli, G. Nenciu, J. Math. Phys. **14**, 1675 (1973).

[26] S. Ciulli, C. Pomponiu, I. Sabba-Ştefãnescu, Physics Reports **17**, 133-224 (1975).

[27] G. Nenciu, Lettere al Nuovo Cimento, **4**, Issue 3, pp. 96-100 (1970).

[28] P. Jizba, T. Arimitsu, Phys. Rev. **E 69**, 026128 (2004).

[29] H. Luschgy and G. Pagès, Electronic Communication in Probability, **13**, 422-434 (2008).

[30] O. V. Besov, V. P. Il'in and S.M. Nikol'skii, *Integral representations of functions and embedding theorems*, Ed. Nauka, pp. 9-40, Moscow (in Russian) (1975).

[31] E. M. Stein, *Singular Integrals and Differentiable Properties of Functions.* Princeton Univ. Press. (1971).

[32] O. Johnson, *Information Theory and the Central Limit Theorem.* Imperial College Press, London (2004).

[33] H. Takayasu, A.-H. Sato, and M. Takayasu, Phys. Rev. Lett. **79**, 966 (1997).

[34] G. Steinbrecher, X. Garbet and B. Weyssow, "*Large time behavior in random multiplicative processes*", arXiv:1007.0952v1, (2010).

[35] G. Steinbrecher, B. Weyssow, Phys. Rev. Lett. **92,** 12503 (2004); T. L. Rhodes et al., Phys. Lett. **A 253**, 181 (1999).

[36] S. Aumaître, F. Pétrélis and K. Mallick, Phys. Rev. Lett. **95**, 064101 (2005).

[37] A. Chicocki, "*Tensor Networks for Big Data Analytics and Large-Scale Optimization Problems*", arXiv:1407.3124v2 [cs.NA] (2014).

[38] M. Vlad, F. Spineanu, J. H. Misguich, and R. Balescu, Phys.Rev. **E 58**, 7359 (1998).

E. Vanden Eijnden and R. Balescu, Phys. Plasmas **3**, 815 (1996).

M. Negrea, I. Petrisor, Phys.Rev. **E 70**, 046409 (2004).

[39] M. Vlad, F. Spineanu, J. H. Misguich, and R. Balescu, Phys.Rev. **E 63**, 066304 (2001).

[40] P. Bak, C. Tang and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).

[41] D. L. Turcotte, Rep. Prog. Phys. **62** 1377–1429 (1999).

[42] P. H. Diamond and T. S. Hahm, Phys. Plasmas **2**, 3640 (1995).

[43] X. Garbet and R. E. Waltz, Phys. Plasmas **5**, 2836, (1998).

[44] S. C. Chapman, R. O. Dendy, and B. Hnat, Phys. Rev. Lett. **86**,2814 (2001).

[45] H. R. Hicks and B. A. Carreras, Phys. Plasmas **8**, 3277, (2001).

[46] L. Garcia, B. A. Carreras, D. E. Newman, Phys. Plasmas, **9**, 841, (2002).

[47] G. Dif-Pradalier, P. H. Diamond, V. Grandgirard, Y. Sarazin, J. Abiteboul, X. Garbet, Ph. Ghendrih, A. Strugarek, S. Ku, C. S. Chang

Phys. Rev. **E 82**, 025401, (2010).

[48] P. Jizba, T. Arimitsu, Annals of Physics, **312**,17–59 (2004).

[49] Rolf Nevanlinna, *Analytic Functions*,Springer Berlin Heidelberg, pp. 26-45 (1970).

[50] B. Oksendal, *Stochastic Differential Equations*, Springer, Heidelberg, N.Y., p. 117 (2000).